You might find this additional information useful...

Supplemental material for this article can be found at:
http://jn.physiology.org/cgi/content/full/jn.00032.2010/DC1

This article cites 83 articles, 19 of which you can access free at:
http://jn.physiology.org/cgi/content/full/104/2/1177#BIBL

Updated information and services including high-resolution figures, can be found at:
http://jn.physiology.org/cgi/content/full/104/2/1177

Additional material and information about *Journal of Neurophysiology* can be found at:
http://www.the-aps.org/publications/jn

This information is current as of August 14, 2010 .

New Method for fMRI Investigations of Language: Defining ROIs Functionally in Individual Subjects

Evelina Fedorenko, Po-Jang Hsieh, Alfonso Nieto-Castañón, Susan Whitfield-Gabrieli, and Nancy Kanwisher
McGovern Institute for Brain Research, MIT, Cambridge, Massachusetts

Fedorenko E, Hsieh PJ, Nieto-Castañón A, Whitfield-Gabrieli S, Kanwisher N. New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *J Neurophysiol* 104: 1177–1194, 2010. First published April 21, 2010; doi:10.1152/jn.00032.2010. Previous neuroimaging research has identified a number of brain regions sensitive to different aspects of linguistic processing, but precise functional characterization of these regions has proven challenging. We hypothesize that clearer functional specificity may emerge if candidate language-sensitive regions are identified functionally within each subject individually, a method that has revealed striking functional specificity in visual cortex but that has rarely been applied to neuroimaging studies of language. This method enables pooling of data from corresponding functional regions across subjects rather than from corresponding locations in stereotaxic space (which may differ functionally because of the anatomical variability across subjects). However, it is far from obvious a priori that this method will work as it requires that multiple stringent conditions be met. Specifically, candidate language-sensitive brain regions must be identifiable functionally within individual subjects in a short scan, must be replicable within subjects and have clear correspondence across subjects, and must manifest key signatures of language processing (e.g., a higher response to sentences than nonword strings, whether visual or auditory). We show here that this method does indeed work: we identify 13 candidate language-sensitive regions that meet these criteria, each present in ≥80% of subjects individually. The selectivity of these regions is stronger using our method than when standard group analyses are conducted on the same data, suggesting that the future application of this method may reveal clearer functional specificity than has been evident in prior neuroimaging research on language.

INTRODUCTION

Three major questions drive research on the neural basis of language. First, what brain regions are involved? Second, are any of these regions specialized for particular aspects of linguistic processing (e.g., phonological, lexico-semantic, or structural processing)? Third, are any of these regions specific to language? Whereas previous neuroimaging research has identified a large number of brain regions sensitive to different aspects of language, precise functional characterization of these regions has proven challenging. Here we consider the possibility that a clearer picture of the functional organization of the language system may emerge if candidate language-sensitive regions were identified functionally within each subject individually.

Prior neuroimaging results do not consistently support functional specificity of brain regions implicated in language. For example, the triangular/opercular parts of the left inferior frontal gyrus (IFG) have been implicated in syntactic process-

ing (e.g., Ben-Schahar et al. 2003; Dapretto and Bookheimer 1999; Santi and Grodzinsky 2007; Stromswold et al. 1996), but other studies have implicated these regions in lexico-semantic processing (e.g., Hagoort et al. 2004, 2010; Rodd et al. 2005) and phonological processing (e.g., Blumstein et al. 2005; Myers et al. 2009). Similarly, anterior temporal regions have been implicated in storing amodal semantic representations (e.g., Patterson et al. 2007), but other studies have implicated these regions in structural processing (Noppeney and Price 2004) or in constructing sentential meanings (Vandenberghe et al. 2002). Similar controversies surround the orbital portions of the left IFG, regions in the left superior and middle temporal gyri and other language-sensitive regions. Furthermore, many of these regions have also been implicated in nonlinguistic cognitive processes, such as music (e.g., Levitin and Menon 2003), arithmetic processing (e.g., Dehaene et al. 1999), general working memory (e.g., Owen et al. 2005), and action representation (e.g., Fadiga et al. 2009).[1] However, the lack of strong consistent evidence from neuroimaging for specificity within the language system, or between linguistic and nonlinguistic functions (see Blumstein 2009 for a recent overview), appears to conflict with decades of research on patients with focal brain damage where numerous cases of highly selective linguistic deficits have been described (e.g., Coltheart and Caramazza 2006) as well as with research using methodologies, such as event-related potentials (ERPs), where some functional dissociations–such as the dissociation between semantic and syntactic processes—are well established (e.g., Friederici et al. 1993; Hagoort et al. 1993; Kutas and Hillyard 1980, 1983; Osterhout and Holcomb 1992; Van Petten and Kutas 1990; see Kaan 2009 for a recent review of the relevant literature; cf. Bornkessel-Schlesewsky and Schlesewsky 2008 for a recent reinterpretation of semantic/syntactic ERP components).[2]

---

[1] Some researchers have articulated proposals that aim to unify the seemingly discrepant findings by postulating subdivisions within the larger functional areas (e.g., Buckner et al. 1995; Dapretto and Bookheimer 1999; Fiez 1997; Heim et al. 2009) or by proposing broader functions for some brain regions [e.g., Hagoort et al.'s semantic unification proposal (Hagoort et al. 2009) or Thompson-Schill and colleagues' cognitive control proposal (Novick et al. 2005; Snyder et al. 2007; Thompson-Schill 2005)].

[2] It is worth noting, however, that low-spatial-resolution methods, like event-related potentials (ERPs), are limited in their ability to provide conclusive answers with respect to the functional specificity of particular cortical regions because similar components could be generated by distinct brain regions, and different components could be generated by the same brain region. Questions of whether a particular brain region is specialized for a particular kind of cognitive process are therefore best answered with high-spatial-resolution methods, such as functional magnetic resonance imaging (fMRI), transcranial magnetic stimulation, intra-operative stimulation and, in some cases, lesion studies.

Address for reprint requests and other correspondence: E. Fedorenko, MIT 46-4141C, Cambridge, MA 02139 (E-mail: evelina9@mit.edu).

One possible reason why neuroimaging studies have found little consistent evidence for functional specificity of language-sensitive brain regions is that virtually all prior studies have relied on traditional group analyses, which may underestimate specificity.[3] Functional regions of interest (fROIs) defined within individual subjects can reveal greater functional specificity by enabling us to pool data from corresponding functional regions across subjects rather than from corresponding locations in stereotaxic space that may differ functionally because of intersubject anatomical variability. However, there is no guarantee that this method will work for language, as it requires a robust localizer that reliably selects the same regions in a short functional scan for each subject, a principled method for deciding which functionally activated regions correspond across subjects, replicability of response in these regions within subjects and between subject groups, and a demonstration that these regions exhibit functional properties characteristic of language regions: strong selectivity for linguistic materials independent of presentation modality. Here we develop a localizer that meets all these criteria.

METHODS

In three experiments, we present a functional localizer for regions sensitive to word- and sentence-level meaning, and we test the robustness and reliability of this localizer. In deciding on a localizer task for high-level cognitive domains like language, there is a trade-off between tasks that produce sufficiently robust activations to be detectable in individual participants and tasks that are relatively selective in targeting a particular cognitive process or type of representation. Language tasks that have been previously shown to produce robust activations in individual participants—used primarily in the clinical literature (cf. Pinel et al. 2007) in the attempt to develop a way to preoperatively localize language-sensitive cortex (e.g., Bookheimer et al. 1997; Petrovich Brennan et al. 2007; Ramsey et al. 2001; Xiong et al. 2000)—have typically involved large contrasts (e.g., reading short texts/generating verbs in response to pictures of objects/covert naming of pictures versus looking at a fixation cross). Although these contrasts do produce robust activations and likely include many of the language-sensitive regions, they also plausibly include brain regions supporting nonlinguistic processing (see Fedorenko and Kanwisher 2009 for further discussion). As a result, we decided to focus on a more selective contrast aimed at targeting regions sensitive to word- and sentence-level meaning. In particular, our main contrast is between sentences and lists of pronounceable nonwords.

This and similar contrasts have been used in many previous studies (e.g., Cutting et al. 2006; Friederici et al. 2000; Hagoort et al. 1999; Heim et al. 2005; Humphries et al. 2006, 2007; Indefrey et al. 2001; Mazoyer et al. 1993; Petersen et al. 1990; Vandenberghe et al. 2002). The goal of the current work is not to demonstrate that some brain regions respond more strongly to sentences than to nonwords (we *expect* some regions to show this response profile based on previous research) but rather to use this contrast to identify a set of brain regions engaged in word- and/or sentence-level processing in individual brains. This ability to identify language-sensitive regions in individual brains enables subsequent investigations of these regions. In particular, once a robust localizer has been developed that reliably identifies a set of language-sensitive regions in individual subjects in

a tractably short scan, the functional profiles of these regions can be investigated in detail by examining these regions' responses to various linguistic and nonlinguistic stimuli. The ultimate goal is to understand the nature of the computations carried out by each of these regions.

It is worth noting that there have been a few previous attempts to define language-sensitive regions functionally using the group data (e.g., Kuperberg et al. 2003) as well as use individual subjects' activation maps in investigations of the neural basis of language (e.g., Ben-Shachar et al. 2004; January et al. 2009; Neville et al. 1998; Pinel and Dehaene 2009). Group-based ROI-level analyses suffer from similar issues as group-based voxel-level analyses because intersubject variability in the locations of activations is likely to lead to averaging across active and nonactive voxels for each subject within the ROI(s), thereby resulting in decreased sensitivity and selectivity (see RESULTS) (see also Fedorenko, Nieto-Castañón, and Kanwisher, unpublished observations, for some evidence; see Nieto-Castañón, Fedorenko, and Kanwisher, unpublished observations, for a detailed discussion). Furthermore none of these studies validated a localizer task independently prior to applying it to the critical question of interest. Instead typically, all or a subset of experimental conditions were used to restrict the selection of voxels for analysis using either the group data (e.g., Kuperberg et al. 2003) or individual subjects' activation maps (e.g., Ben-Shachar et al. 2004; January et al. 2009). [The only exception, to the best of our knowledge, is the work of Pinel et al. (2007) who provide some information on the reliability of a language localizer contrast. However, the goal of the research described in Pinel et al.'s paper is quite different from that of the current work. In particular, our goal is to find a way to reliably identify language-sensitive regions (supporting high-level linguistic processes) at the individual-subject level so that each of these regions can then be carefully characterized in terms of its functional response profile in future studies. To achieve this goal, we are trying to *circumvent* intersubject variability by defining the ROIs in each subject individually and then averaging the responses from the corresponding functional regions across subjects. In contrast, the goal of Pinel et al.'s work is to *explore* the intersubject variability in functional activations by relating it to behavioral and genetic characteristics of the subjects.]

There are two potential problems with these earlier attempts to use individual subjects' data. First, using the same data for defining the ROIs and for extracting the responses will bias the results (e.g., Kriegeskorte et al. 2009; Vul and Kanwisher 2010). And second, it is difficult to interpret the responses of the ROIs to the critical conditions of interest because of the lack of an independent validation of the functional localizer contrasts (e.g., showing that the contrast picks out the same regions reliably within and across subjects and that regions selected by the relevant contrast exhibit some key properties of high-level language-sensitive regions). For the individual-subjects functional localization approach to be effective, what is needed is not ad hoc localizers that are different for each study but a standardized one (or several standardized ones targeting different aspects of language) that has been shown to be effective (see also *Advantages and limitations of using subject-specific functional ROIs* and DISCUSSION). This is exactly what we are trying to establish in the current work.

The three experiments reported here are similar in their design. Therefore we will discuss them jointly throughout METHODS and RESULTS.

Design (experiments 1–3)

All three experiments include four conditions in a blocked design: sentences, word lists (scrambled sentences), jabberwocky sentences (sentences where all the content words are replaced by pronounceable nonwords, like, for example, "florped" or "blay"), and nonword lists (scrambled jabberwocky sentences).[4] The sentences and the nonwords

---

[3] Of course, other factors plausibly contribute to the lack of a clearer picture in the neuroimaging studies of language. These factors include, among others, differences in the functional contrasts used to isolate different components of the language system, differences in the scanning procedures and in the analyses, and an almost nonexistent practice of replicating findings within and across labs (see Fedorenko, Nieto-Castañón and Kanwisher, unpublished observations, for additional discussion).

[4] The motivation for creating word- and nonword-list conditions by scrambling the sentences and the jabberwocky conditions, respectively, was to minimize between-condition differences that were not intended to be a part of our manipulations. For example, omitting function words from the word-lists

conditions are our critical conditions that are used as the localizer contrast. We will therefore focus on these conditions throughout the manuscript. We reasoned that this contrast would be likely to produce sufficiently robust activations to be detectable in individual participants (see e.g., Schosser et al. 1998 for a similar contrast eliciting activations in individual participants; see also appendix d for a brief discussion of why this, and not one of the functionally narrower contrasts, was chosen). We included the two additional conditions (words and jabberwocky) in the preliminary attempt to investigate whether any of the regions we identify specialize for, or at least are more strongly engaged in, lexical-level versus structural aspects of language. We present data from these "intermediate" conditions in appendix d, for completeness (see also Fedorenko, Nieto-Castañón, and Kanwisher, unpublished observations, for some additional results from these conditions), but we will not focus on these conditions because the primary goal of this paper is to describe and validate our localizer task.

The *sentences condition* engages accessing meanings of individual words (lexical processing) and figuring out how the words relate to one another (structural processing), subsequently combining word-level meanings into larger, phrase-level meanings to derive the meaning of the entire sentence. The *nonwords condition* engages neither lexical nor structural processing.[5] Rather it only involves accessing the phonological forms of nonwords and the function words. The contrast between sentences and nonwords is therefore aimed at identifying regions sensitive to word- and sentence-level meaning. Two sample items for each condition (from *experiment 2*) are shown below. (See appendix A for details on the process of constructing the materials used in *experiments 1–3;* this process ensured that low-level properties of the stimuli were matched across conditions; see appendix d for sample items for the words and the jabberwocky conditions.)

Sentences condition sample items: THE DOG CHASED THE CAT ALL DAY LONG and THE CLOSEST PARKING SPOT IS THREE BLOCKS AWAY; nonwords condition sample items: BOKER DESH HE THE DRILES LER CICE FRISTY'S and CRON DACTOR DID MAMP FAMBED BLALK THE MALVITE.

It is important to note that our main contrast (sentences minus pronounceable nonwords) should not include much phono- or morphological processing, as the relevant properties are matched across conditions (see appendix a for details). This contrast is also not designed to target prosodic, pragmatic, or discourse-level processes. Therefore although some of these processes may be included, the questions of the functional specificity of regions supporting these processes would have to be addressed by different functional localizers specifically targeting these other linguistic processes.

In *experiment 1,* in an effort to simulate the most natural reading context, subjects were instructed to simply read each word/nonword silently as they appear on the screen, without making an overt response.[6] *Experiment 2* was conducted to rule out a potential atten-

tional confound in *experiment 1*. In particular, the sentences condition may be more interesting/engaging than the nonwords condition, so some of the activations observed in *experiment 1* may be due to general cognitive alertness. In *experiment 2,* we therefore included a memory task where participants were asked to decide whether a probe word/nonword—presented at the end of each trial—had appeared in the immediately preceding stimulus. Participants were instructed to press one of two buttons to respond. The difficulty of this memory-probe task is inversely related to the expected attentional engagement across conditions in the passive reading task: less meaningful stimuli are more difficult to encode and subsequently remember (e.g., Potter 1984; Potter et al. 2008), and hence attentional confounds should run in the opposite direction of those in the passive reading task in *experiment 1*. Finally, *experiment 3* was conducted to assess the generalizability of the functional regions identified in *experiments 1* and *2* to a new group of participants and determine which of the regions identified in *experiments 1* and *2* are modality independent. In particular, our localizer task is designed to target brain regions that support relatively high-level linguistic processes (lexical and structural processes). Therefore these regions should show a similar response to linguistic stimuli regardless of whether they are presented visually or auditorily. To test the modality independence of the regions identified in *experiments 1* and *2*, in *experiment 3,* participants were presented with both visual and auditory runs of the localizer task.

Participants (experiments 1–3)

Thirty-seven right-handed participants (26 females) from MIT and the surrounding community were paid for their participation [12 participants (9 females) took part in *experiment 1;* 13 (10 females) in *experiment 2;* and 12 (7 females) in *experiment 3*].[7] All were native speakers of English between the ages of 18 and 40, had normal or corrected-to-normal vision, and were naïve as to the purposes of the study. All participants gave informed consent in accordance with the requirements of Internal Review Board at MIT.

Scanning procedures (experiments 1–3)

Structural and functional data were collected on the whole-body 3 Tesla Siemens Trio scanner at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at MIT. T1-weighted structural images were collected in 128 axial slices with 1.33 mm isotropic voxels (TR = 2,000 ms, TE = 3.39 ms). Functional, blood-oxygenation-level-dependent (BOLD) data were acquired in $3.1 \times 3.1 \times 4$ mm voxels (TR = 2,000 ms, TE = 30 ms) in 32 near-axial slices. The first 4 s of each run were excluded to allow for steady state magnetization.

The scanning session consisted of several (between 6 and 14) functional runs. The target length of each of the three experiments was eight runs (several participants completed an unrelated experiment in the same session).[8] The order of conditions was counterbalanced such that each condition was equally likely to appear in the earlier versus later parts of each functional run and was as likely to follow every other condition as it was to precede it. Experimental blocks were 24 s long,

condition may have resulted in activation of some regions in the sentences > words contrast that are engaged in processing function words rather than in the more high-level structural and sentence-level meaning aspects of language. (See appendix a for details on the process of constructing the materials used in *experiments 1–3.*)

[5] It is possible that the nonwords condition will engage lexical processing to *some* extent due to the presence of function words and due to the fact that some of the nonwords include English morphology. It is also possible that the nonwords condition will engage structural processing to *some* extent due to the fact that the function words and the nonwords may in some cases form part-structures. For example, a determiner "the" followed by a nonword like "malvite," as in the sample item shown in the text, may be construed as a noun phrase even if the surrounding syntactic context does not conform to the rules of the English syntax (as it does in the jabberwocky condition). However, critically, the sentences condition should engage both lexical and structural processing to a *much stronger* extent.

[6] Based on the feedback from the first four participants who noted some difficulty in staying awake/alert with no task (other than the reading task), especially during the later runs, we asked each of the subsequent participants

to press a button after each stimulus (sentence, sequence of words, jabberwocky sentence or sequence of nonwords), when a fixation cross briefly appeared on the screen.

[7] Four additional subjects were scanned on *experiment 1*: two were excluded due to their inability to stay in the scanner for the entire duration of the scanning session and two due to excessive head motion. One additional subject was scanned on *experiment 2* and was excluded due to excessive head motion. Three additional subjects were scanned on *experiment 3*: two were excluded due to excessive head motion and one due to extreme sleepiness throughout the scanning session.

[8] Ten of the 13 participants in *experiment 2* completed an unrelated experiment in the same experimental session. Six of these completed this experiment after the localizer experiment, and four of these completed it before the localizer experiment. This unrelated experiment took ~30–40 min.
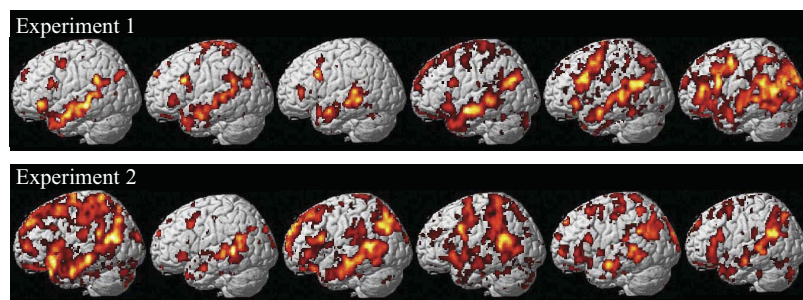
FIG. 1. Sample activations in the left hemispheres of individual subjects for the sentences > nonwords contrast (*top*: sample subjects from *experiment 1; bottom*: sample subjects from *experiment 2*). Threshold: false discovery rate (FDR) 0.05.

with five stimuli (i.e., sentences/word strings/jabberwocky sentences/nonword strings) per block (4 stimuli per block were used in the auditory runs of *experiment 3*), and fixation blocks were 24 s long (*experiment 1*) or 16 s long (*experiments 2* and *3*). Each participant saw between 24 and 32 blocks per condition in a scanning session. (See appendix b for the details of the procedure and timing in *experiments 1–3*.)

General analysis procedures

MRI data were analyzed using SPM5 (//www.fil.ion.ucl.ac.uk/spm) and custom software.[9] Each subject's data were motion corrected and then normalized onto a common brain space (the Montreal Neurological Institute, MNI template). Data were then smoothed using a 4 mm Gaussian filter, and high-pass filtered (at 200 s). Details of the analysis procedures will be discussed in RESULTS.

RESULTS

As discussed in the preceding text, we focus on our main contrast (sentences minus nonwords), which targets regions engaged in retrieving the meanings of individual words (lexical processing) and combining these meanings into more complex—phrase- and sentence-level—representations (structural processing) [see Supplemental Table[10] for a summary of the behavioral results from *experiments 2* and *3;* see appendix d for a discussion of the other 2 conditions built into our design (words and jabberwocky)]. The first three sections test the feasibility of functionally localizing language-sensitive regions in individual subjects. In section 4, we discuss some advantages of the individual-subjects functional localization approach over the traditional group analyses (including a direct comparison between the two methods), as well as potential limitations of this approach.

Activations in individual subjects

At the core of the approach to investigating the neural basis of language developed here is the ability to identify language-sensitive brain regions in individual subjects. We therefore begin by presenting sample activation maps from individual subjects in *experiments 1* and *2* (Fig. 1). As can be seen from Fig. 1, the contrast between sentences and nonwords produces robust and extensive activations in individual brains. Consistent with much of the previous literature examining activations in individual subjects in PET and fMRI studies (for the most part, for preoperative localization purposes), these activations are quite variable, differing in their extent, strength and location (e.g., Bookheimer et al. 1997; Brannen et al. 2001;

Fernandez et al. 2003; Harrington et al. 2006; Herholz et al. 1996; Ramsey et al. 2001; Rutten et al. 2002a,b; Seghier et al. 2004; Stippich et al. 2003; Xiong et al. 2000). Nevertheless, activations appear to land in similar locations across individual brains: for example, in every subject we can see some activations in the left superior and/or middle temporal and temporo-parietal regions, as well as on the lateral surface of the left frontal cortex. Figure 2 further demonstrates that these activations can be obtained in a relatively short period of scanning (~15–20 min) and are replicable within subjects across runs as evidenced by similar activation patterns observed for odd versus even runs (see *Validating a subset of fROIs for further investigation* for a quantitative evaluation of within-subject replicability).

The fact that activations look similar across *experiments 1* and *2* (Fig. 1) demonstrates that our localizer contrast is robust to different tasks (passive reading in *experiment 1* vs. reading with a memory task in *experiment 2*) and materials (as discussed in appendix a completely nonoverlapping sets of stimulus materials were used in the 2 experiments). This generalizability is an important property of a functional localizer as seen for example in localizers for high-level visual areas that are not dependent on particular sets of stimuli or tasks (Berman et al. 2010; Kanwisher et al. 1997).

Now that we have established that the sentences minus nonwords contrast produces robust activations in individual brains in a tractably short period of scanning, we are faced with the challenge of deciding what parts of an individual's activation reflect the activity of distinct regions and how these different parts of the activations correspond across subjects. The traditional method for defining subject-specific fROIs involves examining individual activation maps for the contrast of interest and using
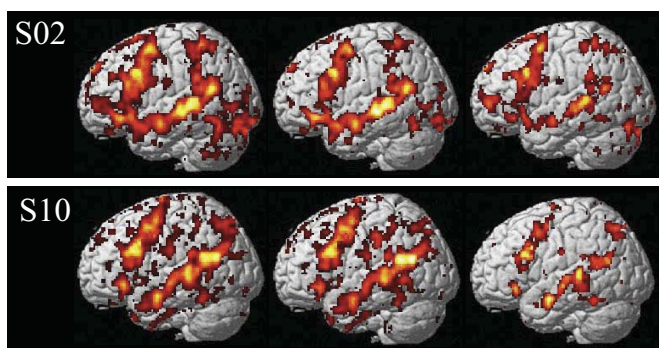

FIG. 2. Activations in 2 sample subjects (*S02, S10*) for all of the runs (*left*: 8 runs in *S02*, 7 runs in *S10*), only the odd-numbered runs (*middle*), and only the even-numbered runs (*right*). [Four runs using the design of the experiments presented here take ~30–40 min; however, this is because these runs include four experimental conditions. With just 2 conditions (sentences and nonwords), which is all that is necessary for functionally defining language-sensitive regions of interest (ROIs), only 2 runs are required, which take ~15–20 min.]

---

[9] The custom scripts developed for the subject-specific analyses used here are publicly available from Fedorenko's website (web.mit.edu/evelina9/www/funcloc.html) in the form of an SPM toolbox, along with the actual localizer task and the group-level partitions discussed in RESULTS.

[10] The online version of this article contains supplemental data.

macro-anatomical landmarks to select the relevant set(s) of voxels (e.g., Kanwisher et al. 1997; Saxe and Kanwisher 2003). This method has worked well in investigating high-level visual regions in the ventral visual pathway (e.g., the fusiform face area, the parahippocampal place area, the extrastriate body area) and also regions implicated in social cognition (e.g., the right and left temporo-parietal junction, the precuneus, the medial prefrontal cortical regions) because these regions are typically located far enough from other regions activated by the target contrast to avoid confusion in defining the regions and in establishing correspondence across subjects. However, as can be seen from Fig. 1, activations for our language contrast are quite distributed and extensive, such that it is difficult to decide on the borders between different parts of the activations as well as on what counts as the "same region" across different brains based on macroanatomy alone. So, a new solution was needed.

New method for defining subject-specific fROIs: group-constrained subject-specific (GcSS) fROIs

Because it is standardly assumed that the same brain regions perform the same cognitive functions across subjects, the process of defining subject-specific fROIs must take into account information about points of high intersubject overlap. The traditional method for identifying spatial overlap across subjects is a whole-brain random-effects group analysis. One possibility for defining *subject-specific* functional ROIs (proposed in Fedorenko and Kanwisher 2009) may therefore involve intersecting the regions that emerge in the traditional group analysis with individual activation maps. We initially tried this method. However, intersecting the random-effects activation map (using data from *experiments 1* and *2*) with the individual activation maps enabled us to capture only 34.8% of individual activations (presumably due to poor alignment of functional activations across brains due to intersubject anatomical variability, which results in missing individual subjects' activations in cases where these activations land in *similar* but mostly nonoverlapping anatomical locations). This seemed low, and so we sought an alternative solution.

The method that we have developed for defining subject-specific fROIs consists of four steps. In *step 1*, a set of individual activation maps for the localizer contrast (sentences minus nonwords, in our case) is overlaid on top of one another. For this step, we included participants from *experiments 1* and *2* ($n = 25$). Because we wanted to later be able to examine the response profile of the resulting regions in an independent subset of the data (e.g., Kriegeskorte et al. 2009; Vul and Kanwisher 2009), we excluded the first functional run in each subject in this step. Each individual subject's activation map was thresholded at $P < 0.05$, corrected for false discovery rate (FDR) (Genovese et al. 2002) for the whole brain volume. The result of this step is a probabilistic overlap map (Fig. 3), showing for each voxel the number of subjects who show a significant sentences > nonwords response at that voxel. Despite similar-looking activations across individual subjects (see Fig. 1), the voxels with the highest intersubject overlap anywhere in the brain have significant activation in only 18 of 25 subjects [there are 2 voxels with this level of intersubject overlap; both are located in the left temporal lobe and have the following coordinates: (1) −56 −44 2; and (2) −62 −42 2].
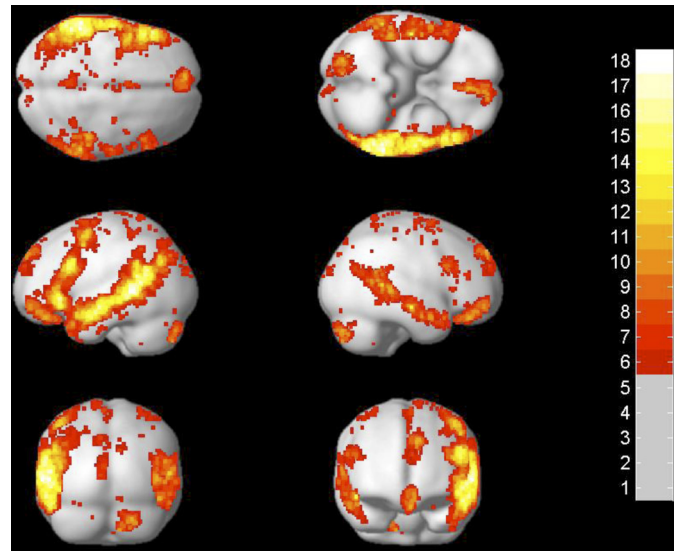

FIG. 3. Probabilistic overlap map for subjects in *experiments 1* and *2* ($n = 25$). Colors indicate the number of subjects showing significant activation for sentences > nonwords in each voxel (the maximum possible value of a voxel equals the number of subjects included in the map, i.e., 25).

This probabilistic overlap map is similar to the random-effects map in that it contains information about points of high intersubject overlap. However, it also contains information about the distribution of individual activations around these high overlap points. Additional advantages to using a probabilistic overlap map instead of a random-effects map for the intersection with individual subjects' activation maps—worth keeping in mind as new functional localizers for language or other domains are developed—are as follows: unlike random-effects maps that are highly dependent on the size of the sample, probabilistic overlap maps are relatively independent of the size of the sample. Furthermore, probabilistic overlap maps are less biased toward high-overlap voxels. However, we would like to stress that the use of a probabilistic overlap map is not the *critical* aspect of these analyses, and in some cases, spatial constraints derived based on a random-effects map (or other kinds of spatial constraints, including anatomical partitions) may be perfectly suitable for constraining the selection of subject-specific voxels (see also discussion in Fedorenko et al., unpublished observations), as long as the partition is large enough to encompass the range of intersubject variability (Nieto-Castañón, Fedorenko, and Kanwisher, unpublished observations). Instead, the critical aspect of our analyses is the intersection with individual activation maps (*step 4* in the following text).

In *step 2*, the overlap map is divided into "group-level partitions" following the topographical information in the map, using an image segmentation algorithm (a watershed algorithm) (Meyer 1991). This algorithm finds local maxima and "grows" regions around these maxima incorporating neighboring voxels, one voxel at a time, in decreasing order of voxel intensity (i.e., number of subjects showing activation at that voxel), and as long as all of the labeled neighbors of a given voxel have the same label. The result of *step 2* is a set of 180 partitions.[11]

---

[11] To minimize the oversegmentation that could result from spurious local maxima, prior to applying the watershed algorithm, we smoothed the overlap map using a 6 mm Gaussian kernel. This step reduced the number of partitions from

The relationship between the size of the partitions and the number of subjects that have a non-zero intersection with the partition
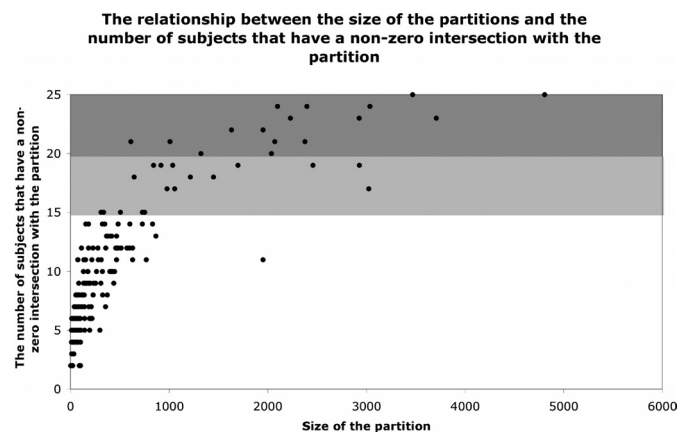
FIG. 4. The relationship between the size of the group-level partitions (for all 180 partitions) and the number of subjects that have a nonzero intersection with the partition (i.e., ≥1 suprathreshold voxel within the borders of the partition). In dark gray are partitions that have a nonzero intersection with ≥80% of the subjects. In light gray are partitions that have a nonzero intersection with 60–79% of the subjects.

In *step 3*, we selected a small subset of these 180 group-level partitions to focus on. Because the overlap map includes voxels present in single subjects, the vast majority of the 180 partitions are small regions specific to individual subjects or small subsets of subjects. The scatter plot in Fig. 4 illustrates the relationship between the size of the partitions and the number of subjects that have a nonzero intersection with them (a nonzero intersection here is defined as a subject having ≥1 above-threshold voxel within the borders of the partition). Because we want to focus on regions that are present in the majority of the subjects, we decided to focus on the subset of the partitions that have a nonzero intersection in ≥20 of 25 (i.e., 80%) of the subjects.[12] There were 16 partitions that satisfied this criterion. The union of these 16 partitions captures a substantial proportion of individual subjects' activations

---

863 (without smoothing) to 180 (with prior 6 mm Gaussian smoothing). In addition, the watershed algorithm was limited to voxels where the smoothed overlap values (a weighted-average of the overlap across neighboring voxels) was ≥1. From the original 137,000 voxels with at least one subject showing activation, 36,323 voxels (27%) had smoothed overlap values <1 (among these, 33,267 voxels had 1-subject, 2,952 voxels had 2-subject, and 104 voxels had 3-subject unsmoothed overlap values). In addition, 11,183 voxels (8%) were determined to lie at the border of two regions (watershed lines) and were subsequently not part of the final partitions (covering 89,494 voxels).

[12] Of course, this cutoff point is arbitrary. One could instead select a larger subset of partitions for further investigation which have a laxer intersection threshold. In fact, although in the current paper we will focus on the regions that have a nonzero intersection with ≥80% of the subjects, on our website (see Footnote 9), we will provide information on an additional set of partitions that intersect with 60-79% of the subjects. Some of these partitions correspond to regions that have been previously implicated in some aspects of linguistic processing, and it may be worthwhile to "track" some of these regions in future studies aimed at characterizing key language-sensitive regions in terms of their role in language processing as well as their specificity for language versus other cognitive processes.

(50%, on average; if we include partitions that overlap with ≥60% of the subjects (a total of 34 partitions), then 64% of individual subjects' activations are captured).

Finally, in *step 4* (the critical step), we defined fROIs in each individual subject by intersecting the chosen subset of group-level partitions with individual activation maps, i.e., the partitions were used to constrain the selection of subject-specific voxels. (No constraints, such as, e.g., contiguity, were placed on the topography of subject-specific voxels within the boundaries of a partition in this step.) Again, we used activation maps for all but the first functional run, leaving out one run to be able to later examine response profiles of these regions using independent data (Kriegeskorte et al. 2009; Vul and Kanwisher 2009). Figure 5 shows the results of this intersection step for a subset of seven sample subjects for two sample partitions: a region in the left inferior frontal gyrus (IFG) and a region in the left middle frontal gyrus (MFG).[13]

In the remainder of the paper, and in future work, we will use the following terminology introduced here: *group-level partitions* are spatial constraints derived based on the probabilistic overlap map (these are intersected with individual subjects' activation maps to define subject-specific fROIs); *GcSS (group-constrained subject-specific) fROIs* will be used to refer to the subsets of subject-specific voxels within the borders of each partition; *group-level fROIs* will be used to refer to fROIs defined based on the traditional group analysis maps (these will only be used in establishing that the new method is at least as good as the traditional methods for detecting functional specificity); and fROIs or simply regions will be used to refer to language-sensitive regions more abstractly, as in, e.g., "the role of the left IFG region in syntactic processing."

Validating a subset of fROIs for further investigation

To establish that this set of brain regions is engaged in language processing, it is necessary to demonstrate that their

---

[13] To evaluate the robustness of our group-level partitions with respect to which particular subset of the data is used in creating the probabilistic overlap map, we performed an additional analysis. In particular, for each subject, we created an additional set of first-level contrasts [e.g., for a subject with 8 runs, contrasts were created for each of the 8 runs as well as for the subsets of runs orthogonal to each of these runs, i.e., runs 2-8 (orthogonal to run 1), runs 1 and 3-8 (orthogonal to run 2), etc.]. (Note that this is a general case of what was done in the analyses reported in the current paper, where only 1 such division was used: run 1 vs. all but the first run.) Then for each subject, we computed the average of the mask files associated with the orthogonal sets of runs. Finally, we overlaid the resulting images across subjects to create a new probabilistic overlap map, which was then divided into partitions using the same watershed algorithm. The partitions which showed high overlap with individual subjects' activation maps were very similar across the two analyses. For example, across our 13 key partitions, mean percentage voxel match—between the two sets—was 88% (SE 1.9). This suggests that our group-level partitions are robust to which particular subset of the data is used in creating the overlap map.
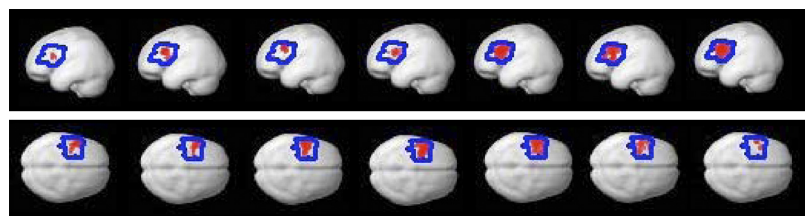

FIG. 5. Two sample functional ROIs [fROIs, left inferior frontal gyrus (IFG) and left middle frontal gyrus (MFG)] in 7 sample subjects. The borders of the group-level partitions are shown in blue and the subject-specific activations are shown in red.

response profile replicates within and between subject groups and generalizes from visual to auditory presentation. We decided on the following four criteria that each fROI must satisfy to be considered a high-level linguistic region.

First, the sentences minus nonwords contrast must be significant in the fROI in a subset of data that was not used to define the fROIs. Second, the sentences condition must be significantly higher in the fROI than the fixation baseline in a subset of data that was not used to define the fROIs. Third, the sentences minus nonwords contrast must be significant in the fROI in a new group of subjects. Fourth, the sentences minus nonwords contrast must be significant in a new group of subjects in the auditory version of the localizer task.

The motivation for the first criterion is straightforward: we are only interested in regions that are replicable within an individual, i.e., that have a stable response profile over time. The current standard in the field for assessing the replicability of a region's response involves defining the region using a subset of the runs and then estimating the response for the conditions of interest in the remaining run(s) (e.g., Krieges-korte et al. 2009; Vul and Kanwisher 2009). The second criterion was included because even if a region shows a highly reliable difference between sentences and nonwords, but the sentences condition elicits a response that is not different from or is below the fixation baseline, it seems problematic to think of such a region as a high-level linguistic region. We therefore chose to focus on regions that respond to the sentences condition reliably higher than the fixation baseline. The inclusion of the third criterion ensured that the regions we identified in *experiments 1* and *2* were not specific to the particular group of subjects tested in those experiments. Finally, the fourth criterion provides a test of an important property that high-level linguistic regions should have: modality independence. In particular, because we are targeting relatively high-level aspects of linguistic processing (processing word- and sentence-level meaning), the modality of the input (visual vs. auditory) should not much affect the response profiles of our regions (cf. Constable et al. 2004; Michael et al. 2001).

The results are summarized in Table 1. Three of the 16 fROIs failed to satisfy at least one of the four criteria and will be excluded from further consideration. Ten of the 16 fROIs satisfied all four criteria. Three additional fROIs satisfied three of the four criteria and showed a marginal effect in the fourth. Given that a conjunction of these four criteria is quite stringent, we will for now include these three ROIs.

Figures 6 and 7 show the group-level partitions corresponding to the 13 fROIs that we selected based on our four criteria (for now, we will refer to them in terms of their approximate anatomical locations relative to macroanatomic landmarks). These regions include four regions in the left frontal lobe (2 regions in the IFG, a region in the middle frontal gyrus, and a region in the superior frontal gyrus), five regions in the left temporal/parietal lobes (4 regions spanning the temporal lobe and a region in the angular gyrus), two regions in the right temporal lobe, and two cerebellar regions (the right and left cerebellar regions are not homologous; we have labeled them right and left cerebellum for now, for simplicity). Figure 8 shows the responses of the GcSS fROIs (defined using these partitions) to sentences and nonwords in an independent subset of the data in *experiments 1* and *2,* in the first visual run in *experiment 3,* and in the auditory runs in *experiment 3* (in

experiment 3 where each participant did several runs of the visual localizer task and several runs of the auditory localizer task, we used all but the first run of the visual data to define subject-specific fROIs and we then extracted the response from the first visual run and from all of the auditory runs).

As can be seen in Fig. 8, every region shows a reliable difference between the sentences and the nonwords conditions in both visual and auditory modalities. Although some regions show main effects of modality, such that the response is overall stronger to visually presented stimuli [e.g., the Left PostTemp, Left AngG or RightCereb regions] or to auditorily presented stimuli [e.g., the Left AntTemp, Left MidAntTemp or Right MidAntTemp regions], we will not explore these modality effects further here, but this is something to keep in mind in future investigations as we gather additional information about the functional profiles of these regions.

It is also worth noting that several of the regions respond to nonwords significantly above baseline, suggesting that these regions may play some role in phonological processing. However, because the nonwords condition includes function words and some of the nonwords include English morphology, this response could also reflect morphological processing and/or processing meanings associated with function words. Future work will be necessary to evaluate the potential role of these regions in phonological and morphological processing.

In addition to the four tests described above, we examined within-subject replicabilty in a different way: we used a multi-voxel pattern analysis (e.g., Haxby et al. 2001) on our 13 group-level partitions. In particular, for each subject (in *experiments 1* and *2*), we split the data into odd- and even-numbered runs and examined the correlations in the activation maps for the sentences minus nonwords contrast—across all voxels in each partition—for odd- versus even-numbered runs *within* each subject versus odd- versus even-numbered runs *between* subjects (each subject's odd runs were compared with every other subject's even runs, and the average value was computed). In Fig. 9 we present the average correlation values for each of our 13 partitions. Within-subject correlations were highly significant (12/13 partitions were significant at $P < 0.0001$; the remaining partition was significant at $P < 0.001$; all P values were FDR-corrected, taking into account all 180 original partitions). Between-subject correlations were significant for all but one partition: 9/13 were significant at $P < 0.0001$, 3/13 were significant at $P < 0.001$. The remaining partition (left Cereb) was marginal at $P = 0.055$. Importantly, however, as Fig. 9 clearly shows, the correlation values are significantly higher for within-subject comparisons (average r value across partitions: 0.52) than for between-subjects comparisons (average r value across partitions: 0.10; paired t-test $P < 0.0001$). These results demonstrate that voxels in each of the partitions show highly replicable responses within subjects. Furthermore, these results underscore the point about intersubject variability in the precise location of functional activations within each partition. As discussed in the preceding text, different subjects show different (sometimes nonoverlapping) subsets of voxels in the sentences > nonwords contrast. As a result, even though a voxel may show a strong and stable effect in one subject, this same voxel may not show an effect in most of the other subjects. This plausibly leads to lower between-subjects correlation values, which are at the heart of the problem with standard group analyses.

TABLE 1. *Basic properties of and a summary of the results for the subset of 16 high-overlap partitions and the corresponding GcSS fROIs*

ROI		Size of the Group-Level Partition	Average Size of a GcSS fROI (Experiments 1 and 2)Parentheses	Peak Coordinate of the Group-Level Partition	S-N Effect Significant in an Independent Subset of the Data	Sentences Significantly Higher Than Fixation	S-N Effect Generalizes to a New Group of Subjects (Visual Presentation)	S-N Effect Generalizes to a New Group of Subjects (Auditory Presentation)
ROIs that satisfied the selection criteria								
Left Frontal ROIs	Left IFG	3468	536 ± 98	−48, 16, 24	<0.0001	<0.0001	<0.0005	<0.0005
	Left IFGorb	2926	400 ± 66	−48, 33, −4	<0.0001	<0.0001	<0.0005	<0.0001
	Left MFG	3036	365 ± 70	−40, −2, 53	<0.0001	<0.0001	<0.05	<0.0005
	Left SFG	2071	234 ± 69	−7, 50, 41	<0.0005	<0.05	0.064	<0.0001
Left Posterior ROIs	Left AntTemp	1952	172 ± 104	−52, 2, −18	<0.0001	<0.0001	<0.0001	<0.0001
	L. MidAntTemp	2099	385 ± 61	−55, −18, −13	<0.0001	<0.0001	<0.01	<0.0001
	L. MidPostTemp	4807	938 ± 112	−56, −40, 10	<0.0001	<0.0001	<0.0001	<0.0001
	Left PostTemp	2397	386 ± 67	−48, −62, 15	<0.0001	<0.0001	<0.0005	<0.0001
	Left AngG	2230	277 ± 65	−37, −76, 30	<0.0001	0.068	<0.005	<0.0005
Right Posterior ROIs	R. MidAntTemp	2038	251 ± 62	55, −14, −13	<0.0001	<0.005	<0.05	<0.0001
	R. MidPostTemp	3709	466 ± 106	58, −45, 10	<0.0001	<0.0001	<0.05	<0.0001
Cerebellar ROIs	Right Cereb	2377	267 ± 69	21, −81, −34	<0.0001	<0.0001	<0.01	<0.0001
	Left Cereb	1323	104 ± 34	−29, −32, −21	<0.0005	<0.01	0.059	<0.0001
ROIs that failed to satisfy the selection criteria								
	Medial PFC1	1008	117 ± 29	−3, 55, −9	<0.01	NS	NS	<0.0005
	Medial PFC2	1632	161 ± 51	−4, 58, 18	<0.01	NS	NS	<0.0001
	Right AntTemp	611	69 ± 17	53, 4, −20	<0.0001	<0.05	NS	<0.005

Values are means ± SE. GcSS, group-constrained subject specific; fROIs, functional regions of interest; IFG, interior frontal gyrus; IFGorb, orbital IFG; MFG, middle frontal gyrus; SFG, superior frontal gyrus; AntTemp, anterior temporal lobe; MidAntTemp, middle-anterior temporal lobe; MidPostTemp, middle-posterior temporal lobe; AngG, angular gyrus; PFC1 and -2, prefrontal cortex.
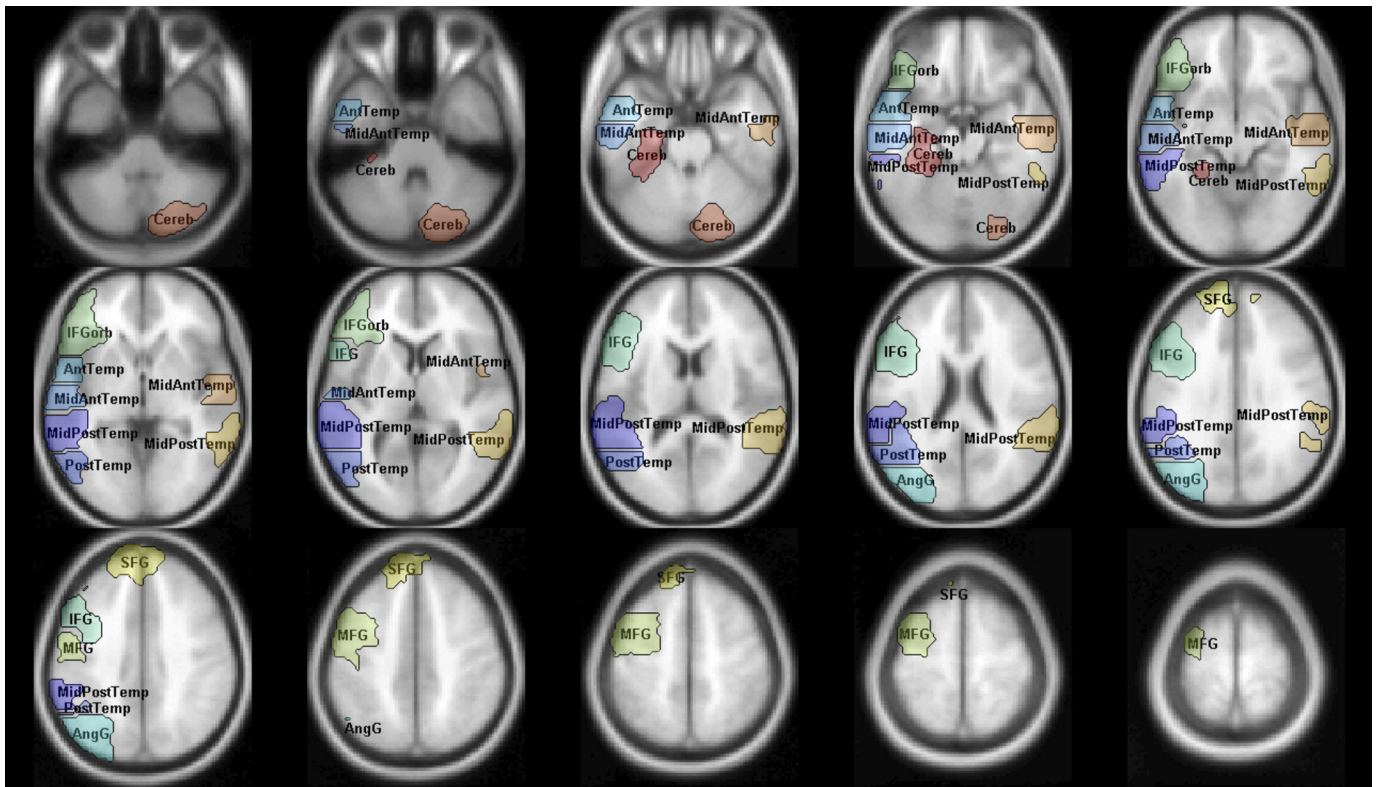
FIG. 6. Group-level partitions corresponding to the 13 key fROIs shown on a slice mosaic.

Advantages and limitations of using subject-specific functional ROIs

ADVANTAGES.   There are several important advantages of the individual-subjects functional localization approach compared with the traditional random-effects group analyses. The most important advantage of this approach is that it allows identification and investigation of the corresponding functional regions across subjects rather than the corresponding locations in stereotaxic space that may differ functionally because of the anatomical variability across subjects. This ability to reliably

and quickly identify a set of key language-sensitive regions in each individual brain enables future studies to investigate the *same functional regions*, allowing us to establish a cumulative research enterprise in our field rather than producing dozens of distinct incommensurable findings. In traditional random-effects group analyses, it is often difficult to relate sets of findings from different studies to one another because it is difficult to evaluate whether activations observed across different studies reflect activity of the "same" functional regions. As a result, new studies do not build on older studies. The
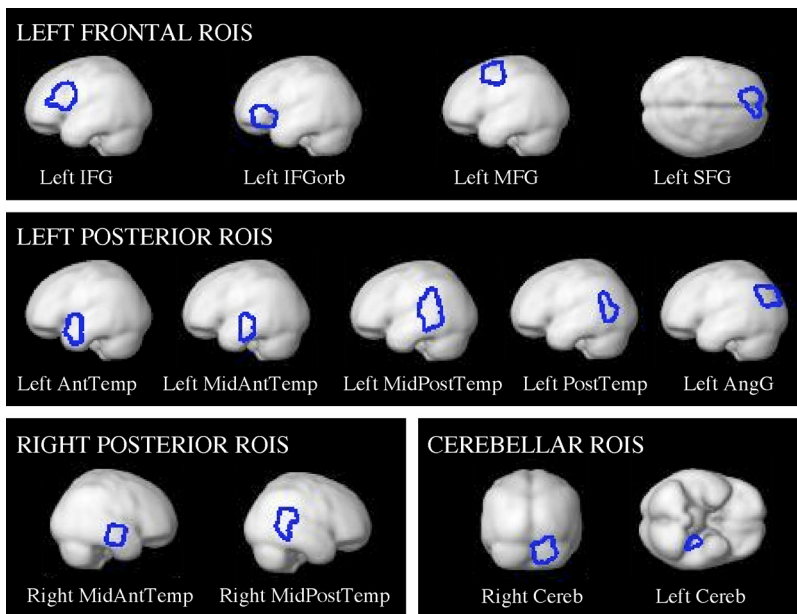

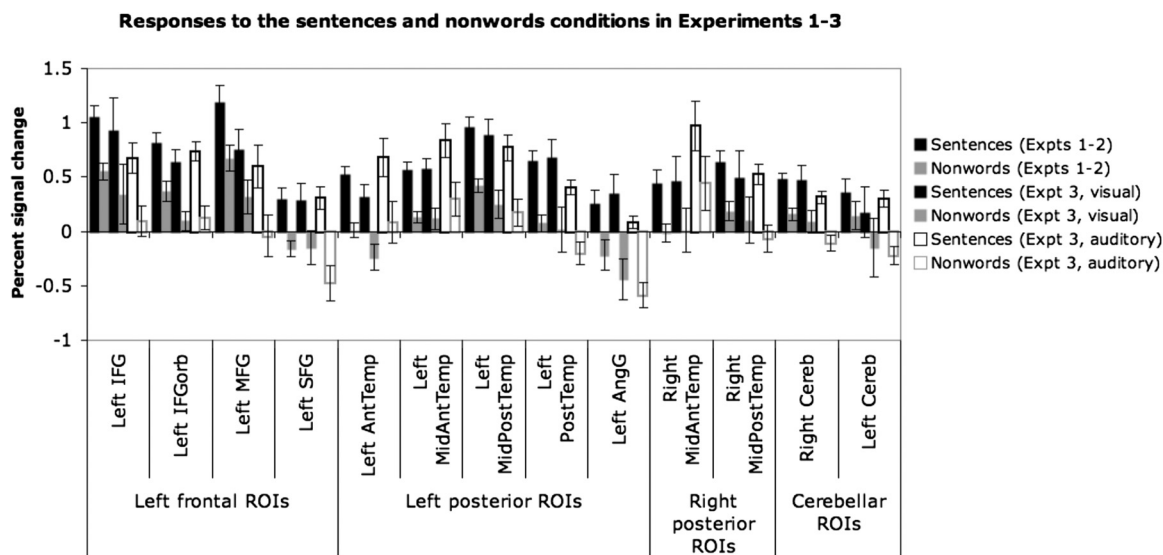FIG. 7. Group-level partitions corresponding to the 13 key fROIs projected onto the brain surface.

FIG. 8. Responses of the 13 fROIs to the sentences and nonwords conditions in an independent subset of the data in *experiments 1* and *2* (*1st 2 bars*), in the 1st visual run in *experiment 3* (*2nd 2 bars*), and in the auditory runs in *experiment 3* (*last 2 bars*). Error bars represent SE.

individual-subjects functional localization approach allows a systematic investigation of a set of language-sensitive regions across studies and across labs, leading to an accumulation of knowledge about functional profiles of these regions (see Fedorenko et al., unpublished observations, for additional discussion). Furthermore, results can be more easily replicated across studies, which should reduce the number of controversies currently present in the field.

Second, because the analyses are restricted to a small number of functionally defined ROIs instead of tens of thousands of voxels, statistical power is much higher in this approach compared with the traditional group analyses (see Saxe et al. 2006 for additional discussion). This higher power allows investigation of potentially subtle aspects of functional profiles of the ROIs.

Third, the ability to identify language-sensitive regions in individual subjects enables investigation of small but scientifically interesting populations, where—in some cases—there may not be a sufficient number of subjects for a traditional group analysis.

Fourth, reliable localizers for key cognitive functions, such as high-level linguistic processing, can be used in preoperative localization in an effort to minimize, and maybe eventually eliminate, the need for electrocortical stimulation mapping, which is highly invasive (e.g., Bookheimer et al. 1997).

Finally, let us briefly consider the two main alternatives to the individual-subjects functional localization approach. The first is the traditional group analysis approach (e.g., a random-effects analysis). As discussed in the preceding text, one potential issue with the traditional group analyses is that they may miss some activations due to insufficient intersubject overlap at the voxel level even though most/all subjects may show activation in/around a particular anatomical location. Another potential problem is the risk of underestimating functional specificity because noncorresponding regions are pooled across subjects (see Nieto-Castañon et al., unpublished observations, for a detailed discussion). To illustrate this point, we performed a direct comparison of functional selectivity (the size of the sentences minus nonwords effect) between our GcSS fROIs and fROIs defined on the group data, i.e., by taking voxels that emerge in a random-effects group analysis and that fall within the borders of our group-level partitions (critically, these group-level ROIs are *identical across subjects*). A priori, one could imagine that the selectivity of the random-effects group-level fROIs would be at least as high as that of GcSS fROIs because the random-effects group-level fROIs include voxels that behave most consistently across participants and may therefore be most functionally selective. However, this is not what we find. As can be seen in Fig. 10, we observe reliably higher selectivity for the subject-specific
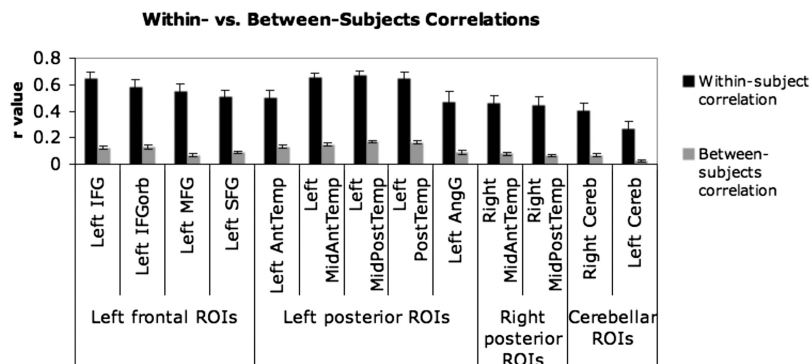

FIG. 9. Correlations across the voxels in each of the 13 group-level partitions comparing odd- vs. even-numbered runs within subjects (■) and comparing odd- vs. even-numbered runs between subjects (▨). Error bars represent standard errors of the mean.

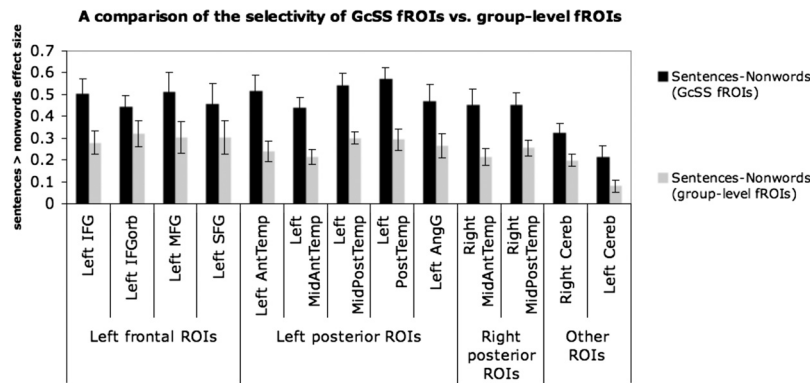**A comparison of the selectivity of GcSS fROIs vs. group-level fROIs**

FIG. 10. A comparison of the selectivity (the size of the sentences > nonwords effect) of group-constrained subject-specific (GcSS) fROIs vs. group-level fROIs (based on the random-effects group analysis of subjects in *experiments 1* and *2, n* = 25). Significance levels: left orbital IFG (IFGorb) < 0.05; left cerebellum < 0.01; left angular gyrus (AngG), left middle frontal (MFG), right cerebellum, left superior frontal gyrus (SFG) < 0.005; the rest of the regions <0.001. Error bars represent SE.

fROIs, compared with the group-level fROIs in every one of the 13 regions (this is similar to the result reported for the selectivity of the fusiform face area defined in individual subjects vs. based on the group data) (Saxe et al. 2006).

The second alternative approach uses anatomical regions of interest, such as Amunts et al.'s (1999) probabilistic cytoarchitectonic maps (e.g., Binkofsky et al. 2000; Heim et al. 2005, 2009; Santi and Grodzinsky 2007). Although these maps are valuable in bringing us one step closer to detecting structure beyond the level of macroanatomy in functional neuroimaging, it is important to keep in mind that, like the random-effects group analyses, these maps may also underestimate functional specificity due to variability in the locations of cytoarchitectonic zones across individual brains, especially in higher-order cortices (e.g., Fischl et al. 2008). (See Fedorenko et al., unpublished observations, for additional comparisons between subject-specific fROIs discussed in this paper and probabilistic cytoarchitectonic maps, as well as other kinds of group-level ROIs, demonstrating that GcSS fROIs exhibit overall stronger response to linguistic stimuli as well as higher functional selectivity.)

In summary, the individual-subjects functional localization approach has the potential to complement existing approaches and may reveal stronger functional specificity within the language system than has been evident in prior research.

LIMITATIONS. Perhaps the most common criticism of the individual-subjects functional localization approach is that the localizer task might miss some key regions involved in the cognitive process of interest. Indeed when dealing with complex high-level cognitive processes, it is difficult, perhaps impossible, to devise a localizer contrast that will capture all and only the brain regions supporting the cognitive process of interest (linguistic processing, in our case). We have three responses to this potential concern. First, the localizer contrast presented in the current paper includes the "classical" (left frontal and left temporo-parietal) language regions, as well as some additional regions commonly implicated in linguistic processing. We are therefore satisfied with the coverage of language-sensitive cortex that our localizer contrast provides. A detailed characterization of the functional profiles of each of these regions will be an important step in understanding how language is implemented in the brain even if this set of regions does not include *all* of the brain regions involved in linguistic processing. Second, the individual-subjects functional localization approach should be supplemented by other analysis methods including whole-brain analyses of each subject and stan-

dard group analyses.[14] These other analysis methods may reveal areas of activation outside the borders of the fROIs as well as pick up regularities in activation patterns at the level below the level of fROIs. When either outcome is found, new fROIs should be considered. And third, this concern should be ameliorated if the conclusions drawn from experiments relying on the fROI approach take into account the fact that the localizer task may not include all the brain regions involved in the cognitive process of interest. For example, if we observed that none of our 13 language-sensitive ROIs responded to some nonlanguage task *x*, we would not be able to argue that language and *x* do not share any brain machinery. Instead what we would be able to conclude is that brain regions supporting the processing of word- and sentence-level meaning (i.e., regions that our localizer task targets) do not respond to *x*. (It is worth noting that this latter issue concerning data interpretation applies to traditional group analyses as well.)

Another potential concern about the individual-subjects functional localization approach arises from the ambiguity of a response profile in which a fROI responds similarly to two conditions: X and Y. In particular, such a response profile could reflect true multi-functionality, such that the fROI supports cognitive processes required for performing X and Y; the response of two adjacent but functionally distinct subregions located within the fROI, such that some of these respond to condition X, and some to condition Y;[15] or the response of multiple small (potentially interleaved) functionally distinct subregions, such that some of these respond to condition X, and the rest respond to condition Y. If the second possibility is the case, then performing a whole-brain analysis on the same data should reveal nonhomogeneous activation patterns within a fROI. This is one reason to always perform such an analysis even when the fROI approach is the primary approach. If the third possibility is the case, then we would not be able to distinguish this response from the response of a truly multi-functional region. However, the traditional methods would not fare better in this case. Multi-voxel pattern analysis (MVPA)

---

[14] There is sometimes a misconception that using the individual-subjects functional localization approach somehow prevents the use of the traditional analysis methods (e.g., Friston et al. 2006). However, there is no reason why group analyses and fROI analyses cannot be conducted on the same data. Indeed, this is generally advisable (Saxe et al. 2006).

[15] It is worth noting that structural heterogeneity has been observed for many Brodmann areas, which are often assumed to constitute "natural kinds." This includes, for example, BAs 44 and 45 (see Amunts et al. 1999 for a systematic investigation and a detailed discussion). This raises similar concerns about adopting BAs as the functional units as those raised with respect to functional ROIs.

methods (e.g., Haxby et al. 2001) are helpful in distinguishing true multi-functionality from the alternative whereby a region consists of multiple functionally-distinct sets of voxels. In particular, even if a region's mean response does not distinguish conditions X and Y, the responses across the whole set of voxels in the region may be different across the two conditions. Therefore the traditional approaches as well as the fROI approach may sometimes be supplemented by MVPA (see Peelen and Downing 2007 for a further discussion).

Yet another potential concern arises from the fact that when using the individual-subjects fROI approach we are not investigating the language system as a whole but rather focusing on several of its component parts, studying each in isolation. However, we would argue that this is not a concern. By carefully characterizing each component of the system, we can eventually get a better handle on the system as whole. This is, in fact, an advantage over the traditional group averaging method where accumulating knowledge about the components of the language system has been difficult because of the difficulty of determining whether some part of activation observed in one study is in the "same" functional region as activation observed in another study. Furthermore, using diffusion tensor imaging and resting state correlation methods, we can begin investigating the relationships among our language-sensitive regions.

Finally, the use of the individual-subjects fROI approach is sometimes seen as "giving up" on anatomy. This is not the case. Rather because the current neuroimaging tools are limited to detecting macroanatomic structures, microanatomic regions vary in where they land relative to sulci and gyri across subjects (e.g., Amunts and Willmes 2006; Amunts et al. 1999; Brodmann 1909; Zilles et al. 1997), and it is *micro*antomic rather than *macro*anatomic structures that have been shown to correspond to function in animal studies (e.g., Iwamura et al. 1983; Matelli et al. 1991; Rozzi et al. 2008), we resort to using functional localizers (combined with anatomical constraints) to identify corresponding functional units across subjects. We are keen to explore possible relationships between our functional regions and cyto-/myelo-architecture, but currently available methods do not allow such an exploration. Until this becomes possible, treating our functional regions as "natural kinds"—keeping an open mind for updating their definitions as we learn more about their functional properties—seems at least as reasonable as (and perhaps more so than) assuming functional correspondence across subjects based on locations in stereotaxic space.

DISCUSSION

Encouraged by the well-documented and substantial advantages of the individual-subjects fROI method in studies of human visual cortex, we sought to test whether such a method might be possible for the domain of language. To be effective, this method requires a robust localizer that reliably picks out the same regions in a short functional scan for each subject, a principled method for deciding which functionally activated regions have correspondence across subjects, replicability of response in these regions within and between subject groups, and a demonstration that these regions exhibit functional properties characteristic of language regions: strong selectivity for linguistic materials independent of modality of presentation. Here we demonstrate that such a method is indeed possible by developing a functional localizer for language cortex that satisfies all four criteria.

Specifically, our localizer, which contrasts the response during reading of sentences versus reading of nonword strings, successfully identifies 13 candidate language regions. Each is present in ≥80% of subjects, each is highly replicable within subjects, and each shows the expected functional profile of a high-level language region. Further, the specificity of these regions for language versus nonlanguage stimuli found using our method far surpasses the specificity of functional responses evident when traditional group analyses are applied to the same data. Now that our method has been validated in the present study, it can be applied to the classic questions of the specificity of language regions for particular aspects of language and for linguistic functions as opposed to nonlinguistic ones.

Before concluding with a summary of the theoretical implications of this work, it is worth emphasizing that defining fROIs is an effort to carve nature at its joints, that is, to identify the fundamental components of a system so that each can be characterized independently. It would be unlikely if the fROIs described here were the best possible characterization of the components of the language system. More likely, future research will tell us that some of these fROIs should be abandoned, some should be split into multiple subregions, others should be combined, and yet other new ones (derived from new functional contrasts) should be added. We intend the use of language fROIs to be an organic, iterative process rather than a rigid and fixed one. On the other hand, to be most useful, some balance will have to be achieved between flexibility of fROI definition and consistency across studies and labs, as the latter is necessary if fROIs are to enable the accumulation of knowledge across studies.

From Gall, Flourens, and Broca, to Chomsky, Fodor, and Pinker, language has resided at the epicenter of the 200-yr-old debate about whether the human mind and brain contain "modules" that are specialized for particular cognitive operations. The apparent evidence from brain-damaged patients for brain specializations for language has become clouded over the last 10 years by a lack of parallel evidence from brain imaging. In a recent review of the brain basis of language, Blumstein (2010) argued that "certain areas of the brain that have been associated with language processing appear to be recruited across cognitive domains, suggesting that while language may be functionally special, it draws on at least some neural mechanisms and computational properties shared across other cognitive domains." We have argued previously that this apparent lack of specificity in the brain imaging literature on language processing may arise in part from the nearly exclusive use of data analysis methods that are known to underestimate functional specificity (Fedorenko and Kanwisher 2009). Here we provide a solution to this problem: a method for analyzing brain imaging data in language studies that does not suffer from the loss of resolution that is intrinsic in methods that rely on spatial overlap of activations across subjects.

To the extent that we find similar results with this new method to those reported in the previous literature (i.e., language-sensitive regions supporting different aspects of language and/or nonlinguistic processes) we will be more confident that this multi-functionality is not a methodological artifact of averaging across adjacent functionally distinct regions across subjects. And to the extent that this method reveals a new picture in which at least some of the language-sensitive regions are highly specialized for a particular kind of linguistic computations, it stands a chance of

reconciling the neuroimaging evidence with the evidence from neuropsychology, thereby leading to clearer answers to fundamental questions about the architecture of the language system and its relationship to other cognitive functions.

APPENDIX A

Details of the materials used in experiments 1–3

All the materials are available from Fedorenko's website (http://web.mit.edu/evelina9/www/funcloc).

Experiment 1

The materials were constructed in several steps: creating the sentences condition, the word-lists condition, the jabberwocky condition, and the nonwords-list condition.

SENTENCES CONDITION. One hundred sixty 12-word-long sentences were constructed using a variety of syntactic structures and covering a wide range of topics.

WORD-LISTS CONDITION. The set of 160 sentences was divided into two sets (sentences 1–80 and sentences 81–160), and the words were scrambled across sentences within each set. Scrambling was done this way to separate the content words in the sentences and the words conditions as far from each other as possible: we used sentences 1–80 in the first half of the runs and sentences 81–160 in the second half of the runs; and we used word-list strings created by scrambling sentences 1–80 in the second half of the runs, and word-list strings created by scrambling sentences 81–160 in the first half of the runs.

JABBERWOCKY CONDITION. Each of the 160 sentences had content words (nouns, verbs, adjectives, adverbs, etc.) removed from it, leaving only the syntactic frame intact. The syntactic frame consisted of function words [articles, auxiliaries, complementizers, conjunctions, prepositions, verb particles, pronouns, quantifiers (e.g., all, some), question words (e.g., how, what), etc., and functional morphemes (past-tense endings, gerund form endings, 3rd person singular endings, plural endings, possessive endings, etc.). Then the content words were syllabified to create a set of syllables that could be re-combined in new ways to create pronounceable nonwords. For syllables that formed real words of English, a single phoneme was replaced (respecting the phonotactic constraints of English) to turn the syllable into a nonword. The syllables were then recombined to create nonwords matched for length (in syllables) with the "corresponding" word (i.e., the word that would appear in the same syntactic frame in the sentences condition). [A post hoc paired-samples t-test revealed no significant difference between words and nonwords in length measured in the number of letters ($P = 0.11$).] This syllable-recombination procedure was used to minimize low-level differences in the phonological make-up of the words and nonwords [e.g., a post hoc paired-samples t-test revealed no significant difference in letter bigram frequency ($P = 0.29$; bigram frequency estimates were obtained from the WordGen program) (Duyck et al. 2004)]. And finally, the nonwords were inserted into the syntactic frames.

We separated the sentences and the jabberwocky conditions as far from each other as possible: we used sentences 1–80 in the first half of the runs and sentences 81–160 in the second half of the runs; and we used jabberwocky sentences 1–80 in the second half of the runs and jabberwocky sentences 81–160 in the first half of the runs.

NONWORD-LISTS CONDITION. The set of 160 jabberwocky sentences was divided into two sets (jabberwocky sentences 1–80 and jabberwocky sentences 81–160), and the nonwords were scrambled across sentences within each set similar to the words condition.

Experiment 2

For *experiment 2,* a new (nonoverlapping with *experiment 1*) set of 160 sentences was created. Like in *experiment 1,* these sentences used a variety of syntactic structures and covered a wide range of topics. In contrast to the materials used in *experiment 1,* the sentences were eight words long and only included mono- and bisyllabic words. This was done because some of the longer words and especially longer nonwords were difficult to read in the allotted time in *experiment 1* (350 ms), and we wanted to keep the time allocated to each word/nonword constant (rather than varying it depending on the length of the word/nonword). The process of constructing the word-lists, jabberwocky and nonword-lists conditions was similar to that used in *experiment 1.*

As discussed in METHODS, *experiment 2* included a memory task. In particular, participants were asked to decide whether a word/nonword—presented at the end of each string of words/nonwords—appeared in the immediately preceding string. Participants were instructed to press one of two buttons to respond. Memory probes were restricted to content words in the sentences and words conditions. Corresponding nonwords were used in the jabberwocky and nonwords conditions. Memory probes required a yes response on half of the trials and a no response on the other half of the trials. True memory probes were selected such that a word/nonword appearing in different positions in the sentence (e.g., at the beginning of the sentence, in the middle of the sentence, or at the end of the sentence) was approximately equally likely to be tested. False memory probes were created by using words and pronounceable nonwords that did not appear in any of the materials in the experiment.

Experiment 3

The materials from *experiment 2* were used (128 of the 160 sentences/word-lists/jabberwocky/nonword-lists; fewer materials were needed because four, not, five items constituted a block of the same length as the visual presentation). For the auditory runs, the materials were recorded by a native speaker using the Audacity software, freely available at http://audacity.sourceforge.net/. The speaker was instructed to produce the sentences and the jabberwocky sentences with a somewhat exaggerated prosody. These two conditions were recorded in parallel to make the prosodic contours across each pair of a regular sentence and a jabberwocky sentence as similar as possible. The speaker was further instructed to produce the words and the nonwords conditions in a way that would make them sound like a continuous stream of speech (rather than like a list of unconnected words). This was done in an effort to minimize the differences in prosody between the sentences and jabberwocky conditions on one hand, and the words and nonwords conditions on the other hand.

APPENDIX B

Details of the procedure and timing in experiments 1–3

PROCEDURE. For *experiments 1* and *2,* and for the visual runs of *experiment 3,* words/nonwords were presented in the center of the screen one at a time in all capital letters. No punctuation was included in the sentences and jabberwocky conditions to minimize differences between the sentences and jabberwocky conditions on one hand and the word-lists and nonword-lists conditions on the other.

Timing

EXPERIMENT 1. The timing was slightly adjusted after the first four participants based on the feedback from the participants. In particular, the per-word timing was changed from 250 to 350 ms. The two timing schemes are described in the following text.

Timing scheme with 250 ms per word. Each trial (each string of 12 words/nonwords) lasted 3,000 ms (250 ms × 12). There were five strings in each block with 600 ms fixation periods at the beginning of the block

and between each pair of strings. Therefore each block was 18 s long. Each run consisted of 16 blocks, grouped into four sets of 4 blocks with 18 s fixation periods at the beginning of the run and between each set of blocks. Therefore each run was 378 s (6 min 18 s) long. There were eight runs in this experiment. The total duration of the experiment (including the anatomical run, setup, and debriefing) was ~2 h.

Timing scheme with 350 ms per word. Each trial (each string of 12 words/nonwords) lasted 4,200 ms (350 ms × 12). There were five strings in each block with 600 ms fixation periods at the beginning of the block and between each pair of strings. Therefore each block was 24 s long. Each run consisted of 16 blocks, grouped into four sets of 4 blocks with 24 s fixation periods at the beginning of the run and between each set of blocks. Therefore each run was 504 s (8 min 24 s) long. There were eight runs in this experiment. The total duration of the experiment (including the anatomical run, set up and debriefing) was ~2 h.

Eleven of the 12 participants completed all eight runs. The remaining participant completed 7/8 runs.

EXPERIMENT 2. Each trial lasted 4,800 ms and included a string of eight words/nonwords (350 ms × 8), a 300 ms fixation, a memory probe appearing on the screen for 350 ms, a period of 1,000 ms during which participants were instructed to press one of two buttons, and a 350 ms fixation. Participants could respond any time after the memory probe appeared on the screen. There were five trials in each block. Therefore each block was 24 s long. Each run consisted of 16 blocks, grouped into four sets of 4 blocks with 16 s fixation periods at the beginning of the run and between each set of blocks. Therefore each run was 464 s (7 min 44 s) long. There were eight runs in this experiment. The total duration of the experiment (including the anatomical run, setup and debriefing) was ~2 h.

Each participant completed between six and eight runs. Ten of the 13 participants completed an unrelated experiment in the same session (see Footnote 8).

EXPERIMENT 3. The timing for the visual runs of *experiment 3* was identical to that used in *experiment 2*. For the auditory runs, each trial lasted 6,000 ms and included a string of eight words/nonwords (total duration varied between 3,300 and 4300 ms), a 100 ms beep tone indicating the end of the sentence, a memory probe presented auditorily (maximum duration 1,000 ms), and a period (lasting until the end of the trial) during which participants were instructed to press one of two buttons. Participants could respond any time after the onset of the memory probe. There were four trials in each block. Therefore each block was 24 s long. Each run consisted of 16 blocks, grouped into four sets of 4 blocks with 16 s fixation periods at the beginning of the run and between each set of blocks. Therefore each run was 464 s (7 min 44 s) long. There were eight runs in this experiment (four visual and four auditory). The total duration of the experiment (including the anatomical run, setup and debriefing) was ~2 h.

Each participant, except for one, completed four visual and four auditory runs. The remaining participant completed two visual and seven auditory runs.

APPENDIX C

A summary of the behavioral data from *experiments 2* and *3*.

Experiment 2

Sentences, Words, Jabberwocky, Nonwords
Accuracy, 98.4 (0.5), 91.8 (2.4), 88.0 (1.7), 85.5 (2.4)
Reaction time 716 (21), 750 (24), 788 (23), 795 (22)
(SE in parentheses)

Experiment 3: visual runs

Sentences, Words, Jabberwocky, Nonwords
Accuracy, 99.3 (0.4), 96.9 (0.8), 91.3 (1.2), 89.2 (1.4)
Reaction time 676 (23), 707 (23), 760 (25), 764 (25)

Experiment 3: auditory runs

Sentences Words Jabberwocky Nonwords
Accuracy, 98.0 (0.8), 93.3 (1.4), 89.4 (1.4), 82.9 (1.4)
Reaction time, 1096 (36), 1138 (42), 1194 (39), 1211 (36)

APPENDIX D

Discussion of the words and jabberwocky conditions

As discussed in METHODS, in addition to the two main conditions (sentences and nonwords), two additional conditions were included in our experiments: words and jabberwocky. There were two reasons for including these conditions. First, we wanted to use these conditions to ask—very preliminarily—whether any of the regions identified with the sentences > nonwords contrast may specialize for, or at least be more strongly engaged in, lexical versus structural processing. And second, we were considering using functionally narrower contrasts (such as words > nonwords or jabberwocky > nonwords) as additional localizer contrasts that may identify more functionally specialized brain regions. We now discuss the results relevant to each of these goals.

Lexical versus structural processing in regions identified with our main localizer contrast

As discussed in the preceding text, the *sentences condition* engages both lexical and structural processing, and the nonwords condition engages neither (cf. Footnote 5). The *words condition* engages lexical processing, but not structural processing, because the individual lexical items can't be combined into more complex representations. The *jabberwocky condition* engages structural processing[16]—due to the presence of function words and functional morphemes distributed in a way allowed by the rules of English syntax—but not lexical processing, beyond perhaps the most rudimentary process of identifying whether a particular nonword denotes an entity, or describes an action or a property, which can be determined based on the syntactic environment a nonword appears in and, in some cases, based on the morphological properties of the nonword. Using this information, it may also be possible to assign thematic roles (e.g., *agent* or *patient*) to entity-denoting nonwords. (For example, in a jabberwocky string "the blay florped the plonty mogg," we can determine that "the blay" and "the mogg" refer to some entities, "florped" describes some action that "the blay" performed on "the mogg," and "plonty" denotes some property of "mogg.") Sample items for the words and jabberwocky conditions (from *experiment 2*): words condition: BECKY STOP HE THE LEAVES BED LIVE MAXIME'S and; SEEN ASLEEP DID FRED TURNED FLOCK THE MUSTARD; Jabberwocky condition: THE GOU TWUPED THE VAG ALL LUS RALL and THE HEAFEST DRODING DEAK IS RHAPH PHEMES AWAY.

Because the sentences condition includes both lexical and structural processing, the higher response to this condition compared with the nonwords condition in our fROIs could be due to lexical processing, structural processing, or a combination of both lexical and structural processing. By examining the response of our fROIs to the two "intermediate" conditions (words and jabberwocky), we can therefore try to determine what factor is the primary contributor to the sentences minus nonwords effect. In particular, if the effect is primarily driven by the processing of word-level meanings, then the fROI should show

---

[16] It is worth noting that jabberwocky sentences (or "syntactic prose") have been commonly used in investigating aspects of structural processing in both fMRI and ERP studies (e.g., Friederici et al. 2000; Hahne and Jescheniak 2001; Indefrey et al. 2001; Moro et al. 2001; Yamada and Neville 2007). The fact that some of these studies have demonstrated similar structural effects for jabberwocky sentences as for sentences containing regular content words suggests that syntactic information is processed similarly in these two kinds of materials.

a higher response to the sentences and words conditions (which include real content words of English that have lexical and referential semantics), compared with the jabberwocky and nonwords conditions. If, on the other hand, the effect is primarily driven by structural processing, then the fROI should show a higher response to the sentences and jabberwocky conditions (which include English syntax, such that the words/nonwords can be combined into more complex representations), compared with the words and nonwords conditions. If both factors contribute to the effect, then the two intermediate conditions should elicit similar responses, lower than that elicited by the sentences condition and higher than that elicited by the nonwords condition.

Figure D1 presents the responses of our GcSS fROIs to the four conditions of the experiment (based on data from *experiments 1* and *2*). As in the analyses in the preceding text, these responses are extracted from the first functional run of each subject, i.e., using data that were not used in defining the fROIs. The responses to the sentences and nonwords conditions are the same as the first two bars for each region in Fig. 8.

To quantitatively evaluate the response profiles of our fROIs to the four conditions, we conducted a 2 × 2 ANOVA on each region, crossing the presence of lexico-semantic information (present, absent) with the presence of structural information (present, absent). In addition, we performed paired-samples *t*-test on pairs of conditions. The results are summarized in Table D1. In most of the fROIs, the response to the two intermediate conditions falls in between the response to the sentences condition and the nonwords condition (surfacing as 2 main effects in the ANOVAs). This response profile is consistent with these regions being engaged in both processing word-level meanings and combining these meanings into more complex representations. However, it is also possible that this design/these materials are not ideally suited for uncovering a dissociation between the cognitive processes involved in lexical access versus syntactic processing. In particular, the words condition may engage structural processing to some extent, because it contains function words that sometimes may result in structure fragments (e.g., "the leaves" or "seen asleep" in D1). It is also possible that the fact that inferring the basic semantic categories of nonwords in the jabberwocky condition (i.e., object, action, property) is sufficient to engage lexical processing to some extent. As a result, these two conditions may elicit a similar response across different language-sensitive regions, some of which may be specialized for lexico-semantic processing and others—for syntactic processing.

Although, for the reasons discussed in the preceding text, it is difficult to interpret the lack of a clear difference in response to the words versus the jabberwocky conditions as indicating a lack of a dissociation between lexical and structural processing in these regions, these results are informative in at least one way. In particular, given that the sentences condition elicits a reliably higher response than either the words or the jabberwocky condition in most regions, it appears to be the case that unconnected lexical representations or the structural frame without the lexico-semantic content cannot elicit as strong a response in these regions as that elicited by the sentences condition, which includes both lexical meanings and a structural frame that allows combining these lexical meanings into a more complex structural/meaning representation.

Future work will be necessary to answer the question of the relative specializations of our fROIs for particular aspects of linguistic processing definitively. For example, the role each of these regions plays in language processing may be illuminated by examining the response of our fROIs to parametric complexity manipulations involving different aspects of the linguistic signal (e.g., lexical complexity, syntactic complexity, discourse-level complexity, pragmatic complexity), or using naturalistic stimuli (e.g., Wilson et al. 2008; Yarkoni et al. 2008) to determine which fROIs are sensitive to different aspects of the signal unfolding over time.

Using functionally narrower contrasts as additional localizer contrasts

In the early stages of our investigation, we observed that although individual subjects' activations (in whole-brain analyses) for the sentences > nonwords contrast were quite systematic across subjects, the activations for the narrower contrasts (words > nonwords, jabberwocky > nonwords) or for the two contrasts directly contrasting the two "intermediate" conditions (words > jabberwocky, and jabberwocky > words) were substantially weaker and highly variable. Consistent with this observation at the level of individual subjects' data, the results of random-effects group analyses on these contrasts (Fig. D2; data from *experiments 1* and *2, n* = 25) revealed little consistent activity. Although in the words > nonwords and the jabberwocky > nonwords contrasts, we do see some activations within the areas that are activated by our main (sentences > nonwords) contrast, these activations are gone when comparing the words and the jabberwocky directly. [Note that the jabberwocky > words contrast reveals some activations in the vicinity of the visual word-form area (e.g., Baker et al. 2007), perhaps because visual/orthographic processing is more effortful for unfamiliar pronounceable nonwords than for familiar words. However, because these regions do not respond to the sentences condition, they do not constitute high-level linguistic regions that we are focusing on here.] This is consistent with the response profiles of our fROIs (Fig. D1) showing similar responses to the words and the jabberwocky conditions. This strengthens our belief that either our fROIs support both lexical and structural processes, or

TABLE D1. *The results of 2 × 2 ANOVAs and t-tests investigating the response of our fROIs to the words and jabberwocky conditions*

	2 × 2 ANOVA Results			*t*-Test Results				
	ME of Syntactic Processing (SJ > WN)	ME of Lexico-semantic Processing (SW > JN)	Interaction	J > N	S > W	W > N	S > J	W vs. J
Left frontal ROIs								
Left IFG	<0.005	<0.05	NS	<0.001	<0.001	<0.005	<0.05	NS
Left IFGorb	<0.005	NS	NS	<0.001	<0.001	<0.05	NS	<0.05 (diff. = 1.181)
Left MFG	0.079	0.095	NS	<0.001	0.06	<0.001	0.054	NS
Left SFG	<0.005	0.091	NS	<0.005	<0.005	<0.05	NS	0.09
Left posterior ROIs								
Left AntTemp	<0.001	<0.005	NS	<0.001	<0.001	<0.05	<0.001	NS
L. MidAntTemp	<0.001	<0.005	NS	<0.001	<0.001	<0.005	<0.001	NS
L. MidPostTemp	<0.005	<0.005	NS	<0.001	<0.001	<0.001	<0.001	NS
Left PostTemp	<0.005	<0.001	NS	<0.01	<0.001	<0.005	<0.001	NS
Left AngG	0.082	0.072	NS	<0.005	<0.05	<0.01	<0.05	NS
Right posterior ROIs								
R. MidAntTemp	0.055	<0.05	NS	<0.05	<0.001	<0.05	<0.001	NS
R. MidPostTemp	<0.05	<0.05	NS	<0.001	<0.005	<0.01	<0.005	NS
Cerebellar ROIs								
Right Cereb	<0.05	<0.01	NS	<0.001	<0.05	<0.001	<0.05	NS
Left Cereb	NS	NS	NS	0.093	<0.05	NS	<0.05	NS

these conditions are not well suited for dissociating between the two. Future work will tell.

REFERENCES

Amunts K, Willmes K. From intersubject variability in clinical syndromes to anatomical variability. *Brain Lang* 96: 147; discussion 157–1701177–150, 2006.

Amunts K, Schleicher A, Bürgel U, Mohlberg H, Uylings HB, Zilles K. Broca's region revisited: cytoarchitecture and intersubject variability. *J Comp Neurol* 412: 319–341, 1999.

Baker CI, Liu J, Wald LL, Kwong KK, Benner T, Kanwisher N. Visual word processing and experiential origins of functional selectivity in human extrastriate cortex. *Proc Natl Acad Sci USA* 104: 9087–9092, 2007.

Ben-Shachar M, Hendler T, Kahn I, Ben-Bashat D, Grodzinsky Y. The neural reality of syntactic transformations: evidence from functional magnetic resonance imaging. *Psychol Sci* 14: 433–440, 2003.

Ben-Shachar M, Palti D, Grodzinsky Y. Neural correlates of syntactic movement: converging evidence from two fMRI experiments. *Neuroimage* 21: 1320–1336, 2004.

Berman MG, Park J, Gonzalez R, Polk TA, Gehrke A, Knaffla S, Jonides J. Evaluating functional localizers: the case of the FFA. *Neuroimage* 50: 56–71, 2010.

Binkofski F, Amunts K, Stephan KM, Posse S, Schormann T, Freund HJ, Zilles K, Seitz RJ. Broca's region subserves imagery of motion: a combined cytoarchitectonic and fMRI study. *Hum Brain Mapp* 11: 273–285, 2000.

Blumstein. Reflections on the cognitive neuroscience of language. In: *The Cognitive Neurosciences* (4th ed.). Cambridge, MA: MIT Press, 2010, 1235–1240.

Blumstein SE, Myers EB, Rissman J. The perception of voice onset time: an fMRI investigation of phonetic category structure. *J Cogn Neurosci* 17: 1353–1366, 2005.

Bookheimer SY, Zeffiro TA, Blaxton T, Malow BA, Gaillard WD, Sato S, Kufta C, Fedio P, Theodore WH. A direct comparison of PET activation and electrocortical stimulation mapping for language localization. *Neurology* 48: 1056–1065, 1997.

Bornkessel-Schlesewsky I, Schlesewsky M. An alternative perspective on "semantic P600" effects in language comprehension. *Brain Res Rev* 59: 55–73, 2008.

Brannen JH, Badie B, Moritz CH, Quigley M, Meyerand ME, Haughton VM. Reliability of functional MR imaging with word-generation tasks for mapping broca's area. *AJNR Am J Neuroradiol* 22: 1711–1718, 2001.

Brodmann K. *Vergleichende Lokalisationslehre der Großhirnrinde.* Leipzig: Barth, 1909.

Buckner RL, Raichle ME, Petersen SE. Dissociation of human prefrontal cortical areas across different speech production tasks and gender groups. *J Neurophysiol* 74: 2163–2173, 1995.

Coltheart M, Caramazza A. *Cognitive Neuropsychology Twenty Years on Macquarie Monographs in Cognitive Science.* London, UK: Psychology, 2006.

Constable RT, Pugh KR, Berroya E, Mencl WE, Westerveld M, Ni W, Shankweiler D. Sentence complexity and input modality effects in sentence comprehension: an fMRI study. *Neuroimage* 22: 11–21, 2004.

Cutting LE, Clements AM, Courtney S, Rimrodt SL, Schafer JG, Bisesi J, Pekar JJ, Pugeh KR. Differential components of sentence comprehension: beyond single word reading and memory. *Neuroimage* 29: 429–438, 2006.

Dapretto M, Bookheimer SY. Form and conent: dissociating syntax and semantics in sentence comprehension. *Neuron* 24: 427–432, 1999.

Dehaene S, Spelke E, Pinel P, Stanescu R, Tsivkin S. Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science* 284: 970–974, 1999.

Duyck W, Desmet T, Verbeke LP, Brysbaert M. Wordgen: a tool for word selection and nonword generation in Dutch, English, German, and French. *Behav Res Methods Instrum Comput* 36: 488–499, 2004.

Fadiga L, Craighero L, D'Ausilio A. Broca's area in language, action, and music. *Ann NY Acad Sci*, 1169: 448–458, 2009.

Fedorenko E, Kanwisher N. Neuroimaging of language: why hasn't a clearer picture emerged?. *Lang Linguistics Compass* 3: 839–865, 2009.

Fedorenko E, Nieto-Castañón A, Kanwisher N. *Syntactic Processing in the Human Brain: What We Know, What We Don't Know, and a Suggestion for How to Proceed* 2010.

Fernández G, Specht K, Weis S, Tendolkar I, Reuber M, Fell J, Klaver P, Ruhlmann J, Reul J, Elger CE. Intrasubject reproducibility of presurgical language lateralization and mapping using fMRI. *Neurology* 60: 969–975, 2003.

Fiez JA. Phonology, semantics, and the role of the left inferior prefrontal cortex. *Hum Brain Mapp* 5: 79–83, 1997.

Fischl B, Rajendran N, Busa E, Augustinack J, Hinds O, Yeo BT, Mohlberg H, Amunts K, Zilles K. Cortical folding patterns and predicting cytoarchitecture. *Cereb Cortex* 18: 1973–1980, 2008.

Friederici AD, Meyer M, von Cramon DY. Auditory language comprehension: an event-related fMRI study on the processing of syntactic and lexical information. *Brain Lang* 75: 289–300, 2000.

Friederici AD, Pfeifer E, Hahne A. Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Brain Res Cogn Brain Res* 1: 183–192, 1993.

Friston KJ, Rotshtein P, Geng JJ, Sterzer P, Henson RN. A critique of functional localisers. *Neuroimage* 30: 1077–1087, 2006.

Genovese CR, Lazar NA, Nichols T. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15: 870–878, 2002.

Hagoort P, Baggio G, Willems RM. Semantic unification. In. *The New Cognitive Neurosciences* (4th ed.). Cambridge, MA: MIT Press, 2009, p. 819–836.

Hagoort P, Brown C, Groothusen J. The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language Cog Processes* 8: 439–483, 1993.

Hagoort P, Hald L, Bastiaansen M, Petersson KM. Integration of word meaning and world knowledge in language comprehension. *Science* 304: 438–441, 2004.

Hagoort P, Indefrey P, Brown C, Herzog H, Steinmetz H, Seitz RJ. The neural circuitry involved in the reading of German words and pseudowords: a PET study. *J Cogn Neurosci* 11: 383–398, 1999.

Hahne A, Jescheniak JD. What's left if the jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Brain Res Cogn Brain Res* 11: 199–212, 2001.

Harrington GS, Buonocore MH, Farias ST. Intrasubject reproducibility of functional MR imaging activation in language tasks. *AJNR Am J Neuroradiol* 27: 938–944, 2006.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293: 2425–2430, 2001.

Heim S, Alter K, Ischebeck AK, Amunts K, Eickhoff SB, Mohlberg H, Zilles K, von Cramon DY, Friederici AD. The role of the left Brodmann's areas 44 and 45 in reading words and pseudowords. *Brain Res Cogn Brain Res* 25: 982–993, 2005.

Heim S, Eickhoff SB, Amunts K. Different roles of cytoarchitectonic BA 44 and BA 45 in phonological and semantic verbal fluency as revealed by dynamic causal modelling. *Neuroimage* 48: 616–624, 2009.

Herholz K, Thiel A, Wienhard K, Pietrzyk U, Von Stockhausen HM, Karbe H, Kessler J, Bruckbauer T, Halber M, Heiss WD. Individual functional anatomy of verb generation. *Neuroimage* 3: 185–194, 1996.
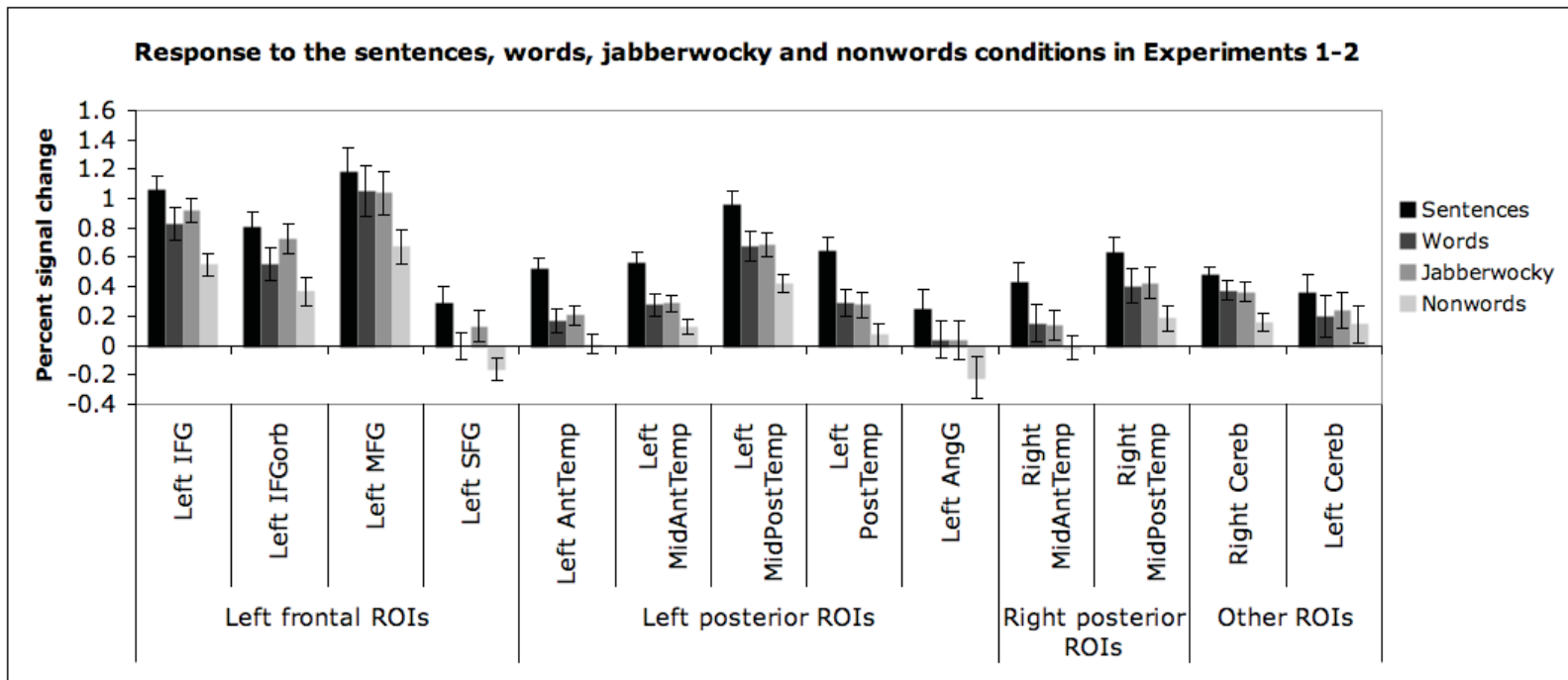
Humphries C, Binder JR, Medler DA, Liebenthal E. Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *J Cogn Neurosci* 18: 665–679, 2006.

Humphries C, Binder JR, Medler DA, Liebenthal E. Time course of semantic processes during sentence comprehension: an fMRI study. *Neuroimage* 36: 924–932, 2007.

Indefrey P, Hagoort P, Herzog H, Seitz RJ, Brown CM. Syntactic processing in left prefrontal cortex is independent of lexical meaning. *Neuroimage* 14: 546–555, 2001.

Iwamura Y, Tanaka M, Sakamoto M, Hikosaka O. Converging patterns of finger representation and complex response properties of neurons in area 1 of the first somatosensory cortex of the conscious monkey. *Exp Brain Res* 51: 327–337, 1983.

January D, Trueswell JC, Thompson-Schill SL. Co-localization of stroop and syntactic ambiguity resolution in Broca's area: implications for the neural basis of sentence processing. *J Cogn Neurosci* 21: 2434–2444, 2009.

Kaan. Event-related potentials and language processing: a brief overview. *Lang Linguistics Compass* 1: 571–591, 2009.

Kanwisher N, McDermott J, Chun, MM. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17: 4302–4311, 1997.

Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat Neurosci* 12: 535–540, 2009.

Kuperberg GR, Holcomb PJ, Sitnikova T, Greve D, Dale AM, Caplan D. Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *J Cogn Neurosci* 15: 272–293, 2003.

Kutas M, Hillyard SA. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207: 203–205, 1980.

Kutas M, Hillyard SA. Event-related brain potentials to grammatical errors and semantic anomalies. *Mem Cognit* 11: 539–550, 1983.

Levitin DJ, Menon V. Musical structure is processed in "language" areas of the brain: a possible role for Brodmann area 47 in temporal coherence. *Neuroimage* 20: 2142–2152, 2003.

Matelli M, Luppino G, Rizzolatti G. Architecture of superior and mesial area 6 and the adjacent cingulate cortex in the macaque monkey. *J Comp Neurol* 311: 445–462, 1991.

Mazoyer, Tzourio, Frak, Syrota, Murayama, Levrier, Salamon G, Dehaene S, Cohen L, Mehler J. The cortical representation of speech. *J Cogn Neurosci* 5: 467–479, 1993.

Meyer. *Un Algorithme Optimal Pour La Ligne De Partage Des Eaux. 8me Congrès de Reconnaissance des Formes et Intelligence Artificielle, Lyon, France,* 1991.

Michael EB, Keller TA, Carpenter PA, Just MA. fMRI investigation of sentence comprehension by eye and by ear: modality fingerprints on cognitive processes. *Hum Brain Mapp* 13: 239–252, 2001.

Moro A, Tettamanti M, Perani D, Donati C, Cappa SF, Fazio F. Syntax and the brain: disentangling grammar by selective anomalies. *Neuroimage* 13: 110–118, 2001.

Myers EB, Blumstein SE, Walsh E, Eliassen J. Inferior frontal regions underlie the perception of phonetic category invariance. *Psychol Sci* 20: 895–903, 2009.

Neville HJ, Bavelier D, Corina D, Rauschecker J, Karni A, Lalwani A, Braun A, Clark V, Jezzard P, Turner R. Cerebral organization for language in deaf and hearing subjects: biological constraints and effects of experience. *Proc Natl Acad Sci USA* 95: 922–929, 1998.

Nieto-Castañón A, Fedorenko E, Kanwisher N. *Subject-Specific Functional Localizers Increase Sensitivity and Selectivity,* 2010.

Noppeney U, Price CJ. An fMRI study of syntactic adaptation. *J Cogn Neurosci* 16: 702–713, 2004.

Novick JM, Trueswell JC, Thompson-Schill SL. Cognitive control and parsing: reexamining the role of Broca's area in sentence comprehension. *Cogn Affect Behav Neurosci* 5: 263–281, 2005.

Osterhout L, Holcomb PJ. Event-related brain potentials elicited by syntactic anomaly. *J Mem Lang* 31: 785–806, 1992.

Owen AM, McMillan KM, Laird AR, Bullmore E. N-Back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Hum Brain Mapp* 25: 46–59, 2005.

Patterson K, Nestor PJ, Rogers TT. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci* 8: 976–987, 2007.

Peelen MV, Downing PE. Using multi-voxel pattern analysis of fMRI data to interpret overlapping functional activations. *Trends Cogn Sci* 11: 4–5, 2007.

Petersen SE, Fox PT, Snyder AZ, Raichle ME. Activation of extrastriate and frontal cortical areas by visual words and word-like stimuli. *Science* 249: 1041–1044, 1990.

Petrovich Brennan NM, Whalen S, de Morales Branco D, O'shea JP, Norton IH, Golby AJ. Object naming is a more sensitive measure of speech localization than number counting: converging evidence from direct cortical stimulation and fMRI. *Neuroimage* 37, *Suppl* 1: S100–108, 2007.

Pinel P, Dehaene S. Beyond hemispheric dominance: brain regions underlying the joint lateralization of language and arithmetic to the left hemisphere. *J Cogn Neurosci* In press.

Pinel P, Thirion B, Meriaux S, Jobert A, Serres J, Le Bihan D, Poline JB, Dehaene S. Fast reproducible identification and large-scale databasing of individual functional cognitive networks. *BMC Neurosci* 8: 91, 2007.

Potter MC. Rapid serial visual presentation (RSVP): a method for studying language processing. In: *New Methods in Reading Comprehension Research* edited by Kieras D and Just M. Hillsdale NJ: Erlbaum, 1984.

Potter MC, Nieuwenstein M, Strohminger N. Whole report versus partial report in RSVP sentences. *J Mem Lang* 58: 907–915, 2008.

Ramsey NF, Sommer IE, Rutten GJ, Kahn RS. Combined analysis of language tasks in fMRI improves assessment of hemispheric dominance for language functions in individual subjects. *Neuroimage* 13: 719–733, 2001.

Rodd JM, Davis MH, Johnsrude IS. The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cereb Cortex* 15: 1261–1269, 2005.

Rozzi S, Ferrari PF, Bonini L, Rizzolatti G, Fogassi L. Functional organization of inferior parietal lobule convexity in the macaque monkey: electrophysiological characterization of motor, sensory and mirror responses and their correlation with cytoarchitectonic areas. *Eur J Neurosci* 28: 1569–1588, 2008.

Rutten GJ, Ramsey NF, van Rijen PC, Noordmans HJ, van Veelen CW. Development of a functional magnetic resonance imaging protocol for intraoperative localization of critical temporoparietal language areas. *Ann Neurol* 51: 350–360, 2002a.

Rutten GJ, Ramsey NF, van Rijen PC, van Veelen CW. Reproducibility of fMRI-determined language lateralization in individual subjects. *Brain Lang* 80: 421–437, 2002b.

Santi A, Grodzinsky Y. Working memory and syntax interact in broca's area. *Neuroimage* 37: 8–17, 2007.

Saxe R, Brett M, Kanwisher N. Divide and conquer: a defense of functional localizers. *Neuroimage* 30: 1088–1096; discussion, 1097–1099, 2006.

Saxe R, Kanwisher N. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage* 19: 1835–1842, 2003.

Schlosser MJ, Aoyagi N, Fulbright RK, Gore JC, McCarthy G. Functional MRI studies of auditory comprehension. *Hum Brain Mapp* 6: 1–13, 1998.

Seghier ML, Lazeyras F, Pegna AJ, Annoni JM, Zimine I, Mayer E, Michel CM, Khatab A. Variability of fMRI activation during a phonological and semantic language task in healthy subjects. *Hum Brain Mapp* 23: 140–155, 2004.

Snyder HR, Feigenson K, Thompson-Schill SL. Prefrontal cortical response to conflict during semantic and phonological tasks. *J Cogn Neurosci* 19: 761–775, 2007.

Stippich C, Mohammed J, Kress B, Hähnel S, Günther J, Konrad F, Sartor K. Robust localization and lateralization of human language function: an optimized clinical functional magnetic resonance imaging protocol. *Neurosci Lett* 346: 109–113, 2003.

Stromswold K, Caplan D, Alpert N, Rauch S. Localization of syntactic comprehension by positron emission tomography. *Brain Lang* 52: 452–473, 1996.

Thompson-Schill SL. Dissecting the language organ: a new look at the role of Broca's area in language processing. *Twenty-First Century Psycholinguistics: Four Cornerstones,* 173–189, 2005.

Van Petten C, Kutas M. Interactions between sentence context and word frequency in event-related brain potentials. *Mem Cogn* 18: 380–393, 1990.

Vandenberghe R, Nobre AC, Price CJ. The response of left temporal cortex to sentences. *J Cogn Neurosci* 14: 550–560, 2002.

Vul E, Kanwisher N. Begging the question: the non-independence error in fMRI data analysis. *Found Philos Neuroimaging* In press.

Wilson SM, Molnar-Szakacs I, Iacoboni M. Beyond superior temporal cortex: intersubject correlations in narrative speech comprehension. *Cereb Cortex* 18: 230–242, 2008.

Xiong J, Rao S, Jerabek P, Zamarripa F, Woldorff M, Lancaster J, Fox PT. Intersubject variability in cortical activations during a complex language task. *Neuroimage* 12: 326–339, 2000.

Yamada Y, Neville HJ. An ERP study of syntactic processing in English and nonsense sentences. *Brain Res* 1130: 167–180, 2007.

Yarkoni T, Speer NK, Balota DA, McAvoy MP, Zacks JM. Pictures of a thousand words: investigating the neural mechanisms of reading with extremely rapid event-related fMRI. *Neuroimage* 42: 973–987, 2008.

Zilles K, Schleicher A, Langemann C, Amunts K, Morosan P, Palomero-Gallagher N., et al. Quantitative analysis of sulci in the human cerebral cortex: Development, regional heterogeneity, gender difference, asymmetry, intersubject variability and cortical architecture. *Hum Brain Mapp* 5: 218–221, 1997.

Supplemental Figure D1 (relevant to Appendix D)

Response to the sentences, words, jabberwocky and nonwords conditions in Experiments 1-2

Supplemental Figure D2 (relevant to Appendix D)

words-
nonwords

jabberwocky-
nonwords

words-
jabberwocky

jabberwocky-
words